Mogul User Guide and Tutorials

2010 CSDS Release

Copyright © 2009 The Cambridge Crystallographic Data Centre
Registered Charity No 800579

Conditions of Use

The Cambridge Structural Database System (CSD System) comprising all or some of the following: ConQuest, Quest, PreQuest, Mercury, (Mercury CSD and Materials module of Mercury), VISTA, Mogul, IsoStar, SuperStar, web accessible CSD tools and services, WebCSD, CSD Java sketcher, CSD data file, CSD-UNITY, CSD-MDL, CSD-SDfile, CSD data updates, sub files derived from the foregoing data files, documentation and command procedures (each individually a Component) is a database and copyright work belonging to the Cambridge Crystallographic Data Centre (CCDC) and its licensors and all rights are protected. Use of the CSD System is permitted solely in accordance with a valid Licence of Access Agreement and all Components included are proprietary. When a Component is supplied independently of the CSD System its use is subject to the conditions of the separate licence. All persons accessing the CSD System or its Components should make themselves aware of the conditions contained in the Licence of Access Agreement or the relevant licence.

In particular:

- The CSD System and its Components are licensed subject to a time limit for use by a specified organisation at a specified location.
- The CSD System and its Components are to be treated as confidential and may NOT be disclosed or re-distributed in any form, in whole or in part, to any third party.
- Software or data derived from or developed using the CSD System may not be distributed without prior written approval of the CCDC. Such prior approval is also needed for joint projects between academic and for-profit organisations involving use of the CSD System.
- The CSD System and its Components may be used for scientific research, including the design of novel compounds. Results may be published in the scientific literature, but each such publication must include an appropriate citation as indicated in the Schedule to the Licence of Access Agreement and on the CCDC website.
- No representations, warranties, or liabilities are expressed or implied in the supply of the CSD System or its Components by CCDC, its servants or agents, except where such exclusion or limitation is prohibited, void or unenforceable under governing law.

Licences may be obtained from:

Cambridge Crystallographic Data Centre
12 Union Road
Cambridge CB2 1EZ
United Kingdom

Web: *http://www.ccdc.cam.ac.uk*
Telephone: +44-1223-336408
Email: *admin@ccdc.cam.ac.uk*

(UNITY is a product of Tripos, L.P. and MDL is a registered trademark of Elsevier MDL)

Mogul User Guide: Graphical Interface

1 Introduction
2 Preparing a Query Structure
3 Importing a Query Structure
4 Editing a Structure
5 Drawing and Editing 2D Structures: The Mogul Drawing Area
6 Running Searches
7 Viewing and Analysing Results
8 Viewing Hit Structures
9 Output Options

1 Introduction

1.1 Overview of Mogul

Mogul is a knowledge base of molecular geometry derived from the Cambridge Structural Database (CSD) and provides rapid access to information on the preferred values of bond lengths, valence angles and acyclic torsion angles; and the preferred geometry of isolated ring systems.

A complete molecule or ion, or a crystal structure containing several complete molecules or ions, should be input to Mogul. A range of file formats are accepted. Alternatively, queries can be drawn using a sketching tool in the Mogul graphical user interface. A search can be performed by selecting a bond, valence angle, torsion or ring in the query molecule. Mogul calculates the values of a set of keys that capture atom- and bond-property information and collectively characterise the environment of the selected molecular feature. Traversal of a search tree indexed on these keys is then used to retrieve all entries from the CSD that have the same molecular feature (i.e. bonds, angles, torsionsor rings with the same set of key values). This is roughly equivalent to an exact substructure search but without the need for graph matching.

A histogram and summary statistics of the distribution of the bond length, valence angle, torsion angle or ring geometry in matching CSD entries are displayed in the program interface or written out to a text file.

There are two ways of using Mogul:

- Interactively via the graphical user interface.
- As a background job, using an instruction file, see **Mogul User Guide: Instruction File Interface**. This mechanism allows batch processing of multiple structures and has been used to integrate Mogul with third-party software, e.g. CRYSTALS, a package for single crystal X-ray structure refinement: *http://www.xtl.ox.ac.uk/crystals.html.*

Mogul has many potential applications including:

- Conformation validation, for example of calculated conformations, or for filtering out protein-ligand docking solutions involving unlikely ligand conformations, etc.
- Geometry validation, for example checking the molecular dimensions of new crystal structures.
- Creation of restraint data/ligand dictionaries for protein structure refinement or to guide small-molecule structure solution from powder diffraction data.

1.2 Chemical Coverage: The Mogul Library

There are currently four libraries available with Mogul: bond length, valence angle, acyclic torsion angle and ring conformation. These libraries provide access to all the experimentally-determined values of bond lengths, valence angles, acyclic torsion angles and rings in the CSD, with the following restrictions:

- Large distributions (> 10,000 observations) are reduced to exactly 10,000 observations by random selection. The maximum size for ring distributions is 500 members.
- Bonds, angles and torsions involving hydrogen atoms are not included.
- Valence angles and torsions involving metals are not included in the Mogul libraries. *Note*: Metal-containing bond fragments and rings are characterised in exactly the same way as organic fragments.
- Only acyclic torsion angle data are included.
- Only rings containing five or more atoms that do not form part of a fused or bridged ring system are included.

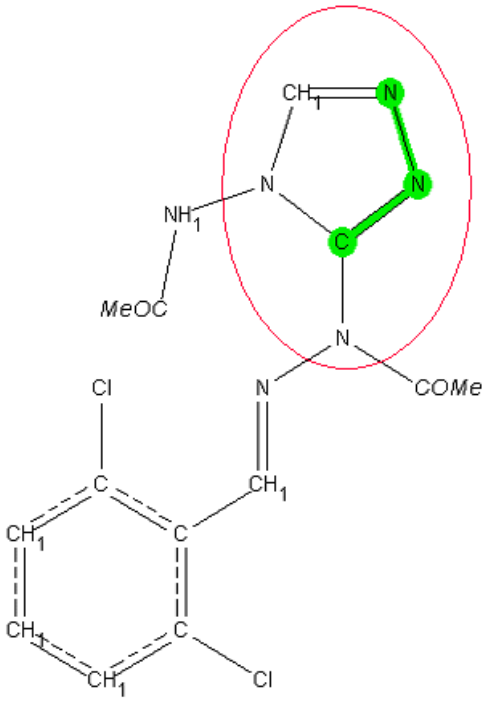
2 Preparing a Query Structure

2.1 Query Structure Preparation

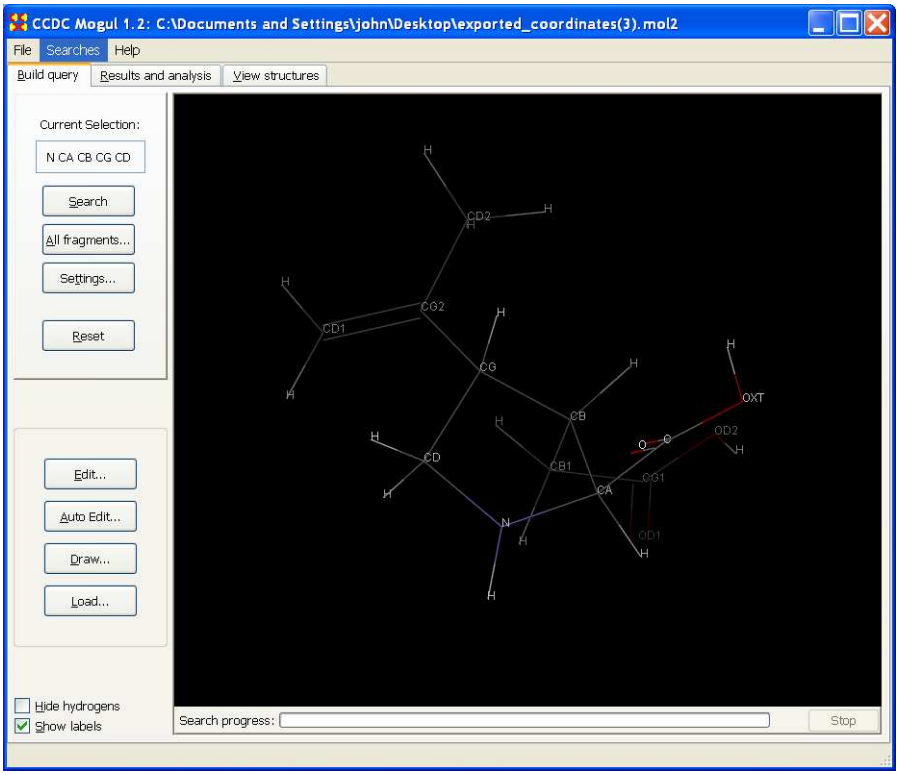
- In order for Mogul searches to run correctly it is essential that query structures are set up properly.
- A query structure can consist of a single molecule, a pair of molecules or ions, or a larger assembly of molecules or ions.
- When preparing query structures it is critical that:
  - Complete molecules or ions are specified (including hydrogen atoms), or at least as much of the structure is present as is required to completely define the chemical environment(s) of the fragment(s) of interest (see Required Molecular Information).
  - Correct conventions for bond types are used (see Bond Type Conventions).
- Query structures can drawn using the Mogul Drawing Area. Alternatively they can imported as 3D sructures. Ring searches can however only be initiated if the model is a 3 dimensional structure as some of the 3dimensional information is required to assess the relevance of a given ring.

2.2 Required Molecular Information

- When Mogul searches for a bond length, valence angle, torsion, a search substructure is generated that extends outwards from the fragment of interest by two bonds in all directions (i.e. the chemical environments of all atoms bonded directly to the search fragment are considered).
- For example, if searching for the C=N-N valence angle in the molecule shown below, all atoms and bonds within the circle will be included in the Mogul search substructure:



- Ring searches retrieve only those rings which have identical Sybyl atom types in the same order as the query ring. Rings are also characterised by the number of substituents attached to each ring atom (usually 0, 1 or 2), the size of the substituent (**Small**: Zero or one heavy atoms attached to the alpha atom of the substituent, **Large**: two or more heavy atoms attached to the alpha atom), and the relative stereochemistry of substituents around the ring (**Up** or **Down**, for single substituents attached to sp3 carbon/nitrogen, **In Plane** for substituents attached to sp2 carbon/nitrogen). Enantiomeric forms of a ring structure are treated such that where complete rings+substituent designations are found that are chiral to each other, these are treated as equivalent and combined in the same distribution.
- For example, the ring in the picture below would be designated as being *Csp3 Csp3 Csp3 Csp3 NSp3* with a single **large** substituent **up** on *CA*, a **small** substituent **down** on *CB* (only one heavy atom off *CB1*) and a **large** substuituent **down** on *CG*. The down and up assignments are important relative to each other only, these substituents could equally well have been assigned **down, up, up**.



- It is essential that the bond types, element types and hydrogen counts within the Mogul search substructure are correct and complete; therefore:
  - Atom and bond types must be defined unambiguously, i.e. queries cannot include variable atom or bond types.
  - All hydrogen atoms must be included. Ensure that any hydrogen atoms added reflect the desired ionisation state of the molecule or ion. For example, the protonation state of a carboxylic acid group can be controlled by adding or removing the ionisable hydrogen atom and this will affect the results.
  - The exact 3D positions of any added hydrogen atoms are not important, but the hydrogen atoms must be bonded to the correct heavy atoms.
  - Atomic charges are ignored during a Mogul search, even if they are set. However, the ionisation state of a molecule or ion will be inferred from the presence or absence of hydrogen atoms.
- If an atom is disordered, only one of its positions should be included in the molecule or ion.
- Query structures will normally consist of one or more complete molecules or ions. However, it is only strictly necessary to input as much of the molecule as is needed to define the environment(s) of the fragment(s) of interest.
- If a query structure is polymeric, such as a catena metal complex, it will not be possible to transfer complete molecules or ions into Mogul. Importing a single monomer unit is unlikely to be sufficient because the chemical environments of molecular features at the edge of the

monomer unit will not be fully defined. You should therefore ensure that enough of the polymeric network is included to completely define the environments of all bond lengths, valence angles and torsions of interest. A dimer unit will normally be sufficient.

2.3 Bond Type Conventions

- Mogul bond types follow the bond type conventions used in the Cambridge Structural Database (CSD) with the exception that polymeric bond types are not used in Mogul. It is critical that these bond type conventions are followed in query structures, otherwise relevant hits may be missed.
- Mogul bond types include:
  - Single
  - Double
  - Triple
  - Quadruple - *occurs for some metal-metal bonds*
  - Aromatic
  - Pi - *used for bonds between metals and pi-bonded (eta-coordinated) ligands*
  - Delocalised - *used for some conjugated systems*
- In general, use of these bond types follows normal chemical principles. However, for some common chemical groups and ring systems there are some arbitrary conventions which must be followed; the more important are tabulated (see **Appendix B: Bond Type Conventions for Common Chemical Groups**).
- Following these guidelines will normally ensure that you obtain reasonable search results. However, due to inconsistencies in the CSD itself, not all entries will follow these conventions rigidly and some hits may be missed.
- Mogul offers options for guessing the bond types of an input query structure, adding missing hydrogen atoms and/or standardising the bond types to CSD conventions (see Assignment of Unknown Bond Types and Missing Hydrogen Atoms). However, these are not guaranteed to work, particularly if the query structure has a very poor geometry.

3 Importing a Query Structure

3.1 Input File Formats

A query structure will typically consist of a single molecule, or multiple molecules or ions. Both 2D and 3D structures can be used (see Loading a Molecule into Mogul). Acceptable file formats include:

- **CIF** (*.*cif*): International Union of Crystallography format for crystal structures (*http://www.iucr.org/iucr-top/cif/home.html*).
- **Mol2** (*.*mol2*): Tripos Inc. format for 3D molecules and crystal structures.
- **PDB** (*.*pdb*): Protein Data Bank format for 3D molecules.
- **SHELX** (*.*res*)**:** Crystal structure file format used by the program SHELX (*http://shelx.uni-ac.gwdg.de/SHELX/*).
- **Mol** (*.mol): Molecular file format produced by MDL Informations Systems Inc.
- **ConQuest Sketcher file** (*.*con*): 2D chemical diagram output from ConQuest.

3.2 Loading a Molecule into Mogul

Input files can be loaded into Mogul in several ways:

- Select **File** from the top-level menu, and **Load Molecule...** from the resulting pull-down menu.
- Click on the **Load** button at the bottom left-hand of the Build query pane.
- If you are using a computer that supports *Drag and Drop*, you can drag a file icon and drop it onto the Mogul program (or a shortcut to it). Mogul will launch and open the dropped file. If Mogul is already open, the file icon can be dragged and dropped into the *Build query* viewing area.
- Alternatively, 2D queries can be constructed using the Mogul *Draw* window (see Drawing and Editing 2D Structures: The Mogul Drawing Area).

Mogul can accept both 2D and 3D input query structures.

Once a query structure has been loaded, Mogul will identify and attempt to assign any unknown bond types, standardise bond types to CSD conventions, and add missing hydrogens (see Assignment of Unknown Bond Types and Missing Hydrogen Atoms).

3.3 Assignment of Unknown Bond Types and Missing Hydrogen Atoms

On loading a structure into the *Build query* pane, Mogul will automatically:

- Deduce the probable bond types of any bonds in the structure whose types are not specified in the input file.
- Standardise all bond types to Cambridge Structural Database (CSD) bond-type conventions (i.e. aromatic and delocalised bond type will be used where appropriate).
- Add hydrogen atoms if none are present in the input file.

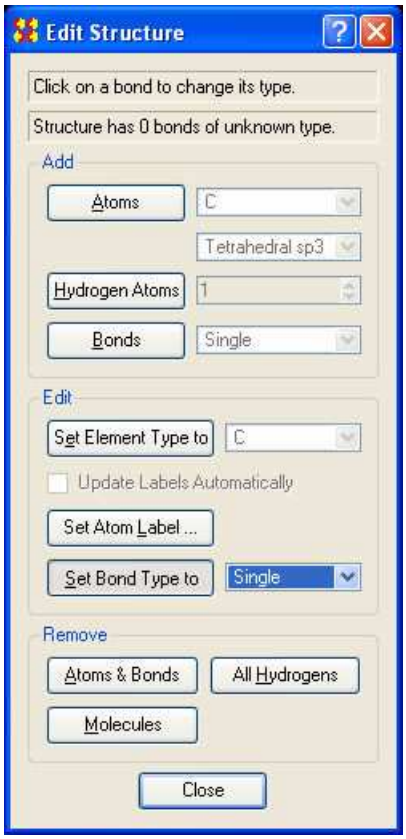
Note: Automatic assignment of unknown bond types is not possible for 2D input queries.

There is no guarantee that bond-type deduction or hydrogen-atom addition will be completely correct. Consequently, you should check the changes that Mogul has made. A summary of the changes made will be given in a *Structure edited* pop-up. Hit **OK** to accept the changes, or **Revert** to reject the changes and return to the original structure.

4 Editing a Structure

4.1 Manually Editing a Structure

To manually edit a query structure within Mogul, select the **Edit...** button to the left of the *Build query* pane. If your structure is 2D this will launch the Mogul *Draw* window (see Drawing and Editing 2D Structures: The Mogul Drawing Area); for 3D query structures the *Edit Structures* window will appear:



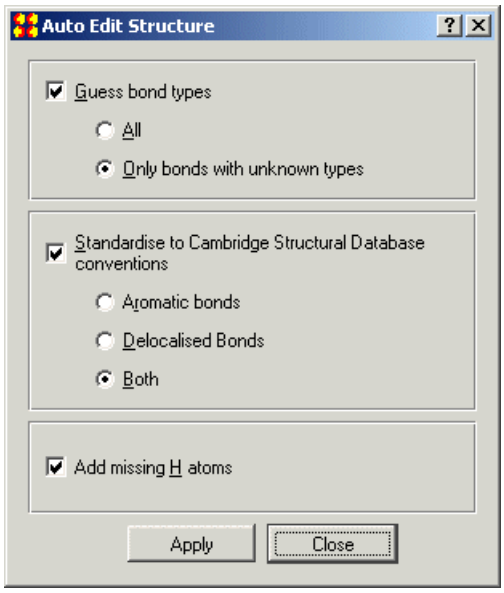
Within this window you can choose to:

- Add atoms: click on the **Atoms** button, select the element type and hybridisation required from the pull down menus, then click on the atom to which the new atom is to be added.
- Add hydrogen atoms: click on the **Hydrogen Atoms** button, select the number of required hydrogen atoms, then click on the atom(s) to which hydrogens are to be added.
- Add bonds: click on the **Bonds** button, choose the bond type from the pull-down menu, then click on the two atoms that you wish to be bonded.
- Edit an element type: click on the **Set Element Type to** button, select the element type required from the pull down menu, then click on the atom whose element type you wish to change. Enable the **Update Labels Automatically** check-box to change the atom label to that chosen.
- Edit an atom label: click on the **Set Atom Label...** button, click on the atom whose label you wish to change in the display, then enter the new label in the resulting *Edit atom* dialogue.

- Edit a bond type: click on the **Set Bond Type to** button, choose the bond type from the pull-down menu, then pick the bond(s) you wish to change.
- Remove atoms and bonds: click on the **Atoms & Bonds** button, then pick on the atoms or bonds you wish to remove.
- Remove all hydrogens: click on the **All Hydrogens** button, then click on an atom or bond in the structure from which you wish to remove all hydrogen atoms.
- Remove complete molecules: click on the **Molecules** button, then pick an atom or bond from within the molecule you wish to remove. It is also possible to only remove a particular copy of a molecule.

4.2 Auto-Editing Options

There are a number of options for editing a query structure automatically; these can be accessed by selecting the **Auto Edit...** button to the left of the *Build query* pane. This will launch an *Auto Edit Structure* window:



The options available within this window are:

- Guess bond types: select the **Guess bond types** check box, then apply this to either **All** or **Only bonds with unknown types**.
  Note: **Guess bond types** is not available for 2D queries.
- Standardise bond types: select the **Standardise to Cambridge Structural Database conventions** check box. This can be applied to **Aromatic bonds**, **Delocalised bonds** or **Both.** It is important that all bond types do follow CSD conventions, otherwise Mogul may fail to find hits (see Bond Type Conventions).
- Add missing hydrogen atoms: select the **Add missing H atoms** check box. Failure to include hydrogen atoms in a query structure may cause Mogul to miss hits (see Required Molecular

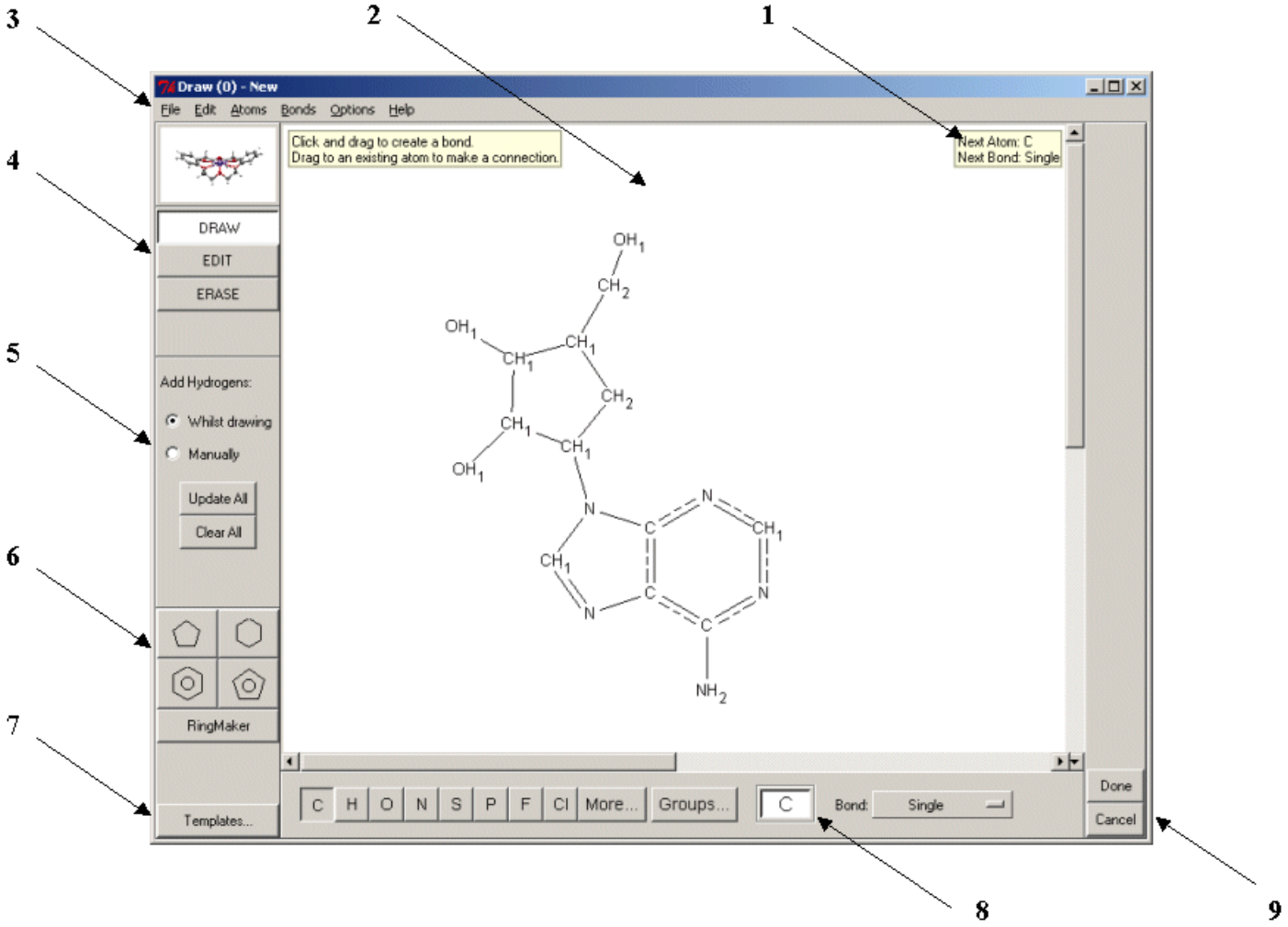
Information).

Hit the **Apply** button to automatically edit your structure according to the selections made.

5 Drawing and Editing 2D Structures: The Mogul Drawing Area

5.1 Layout of the Draw Window

The *Draw* window can be opened by hitting the **Draw button** in the *Build query* pane.



1. Help messages
2. Drawing area (see Fundamentals of Drawing)
3. Top-level menu (different from the menu in the main Mogul interface).
4. Mode buttons (see Modes in the Draw Window) – *responses to mouse clicks in the drawing area will depend on which mode is active*.
5. Hydrogen-atom addition options (see Automatic Addition of Hydrogen Atoms) and (see Adding Hydrogen Atoms Manually).
6. Area for selecting basic ring templates (basic carbon rings to aid drawing) (see Adding a Ring to a Blank Drawing Area)
7. Area for selecting structure templates (molecular building blocks to aid drawing).

8. Area for changing the current element type (see Changing the Current Element Type) and bond type (see Changing the Current Bond Type).
9. Buttons for transferring the drawn structure to the *Build query* pane or cancelling all changes and closing the *Draw* window.

5.2 Modes in the Draw Window

Three mode buttons are available on the left-hand side of the *Draw* window. What happens when the mouse is used in the drawing area will depend on which mode is active. To activate a mode, click on the corresponding button: when active, a mode button is coloured white.

- **DRAW**: Activate this mode to draw a structure.
- **EDIT:** Activate this mode to perform editing tasks such as moving, rotating or resizing substructures, or selecting atoms or bonds.
- **ERASE**: Activate this mode to delete atoms or bonds.

5.3 Fundamentals of Drawing

All drawing takes place in the central white area of the *Draw* window. You can then:

- Draw bonds (see Drawing a Bond).
- Draw isolated atoms (see Drawing an Isolated Atom).
- Draw bonds from existing atoms (see Drawing a Bond from an Existing Atom).
- Draw bonds to existing atoms (see Drawing a Bond to an Existing Atom).
- Draw bonds between existing atoms (see Drawing a Bond between Two Existing Atoms).
- Undo mistakes (see Undoing Mistakes when Drawing Substructures).
- Select atoms and bonds (see Selecting Atoms and Bonds).
- Delete atoms and bonds (see Deleting Atoms and Bonds).
- Perform advanced drawing options such as moving, copying or resizing substructures (see Advanced Drawing Options).

5.3.1 Drawing a Bond

- Ensure you are in **DRAW** mode.
- Move the cursor into the white area of the *Draw* window.
- Press down the left-hand mouse button, move the cursor *while keeping the mouse button depressed*, and then release the button.
- This draws a bond, using the current element type (see Changing the Current Element Type) and bond type (see Changing the Current Bond Type).

5.3.2 Drawing an Isolated Atom

- Ensure you are in **DRAW** mode.
- Move the cursor into the white area of the *Draw* window.
- Click the left-hand mouse button, and release it again without moving the mouse.

5.3.3 Drawing a Bond from an Existing Atom

- Ensure you are in **DRAW** mode.
- Move the cursor onto the atom (the atom will go red).
- Press down the left-hand mouse button.
- Move the cursor while keeping the mouse button depressed, then release the button.

5.3.4 Drawing a Bond to an Existing Atom

- Ensure you are in **DRAW** mode.
- Move the cursor into the white area of the *Draw* window.
- Press down the left-hand mouse button.
- Move the cursor onto the desired atom (the atom will go red) while keeping the mouse button depressed, then release the button.

5.3.5 Drawing a Bond between Two Existing Atoms

- Ensure you are in **DRAW** mode.
- Move the cursor onto the first atom (the atom will go red).
- Press down the left-hand mouse button.
- Move the cursor onto the second atom (the atom will go red) while keeping the button depressed, then release the button.

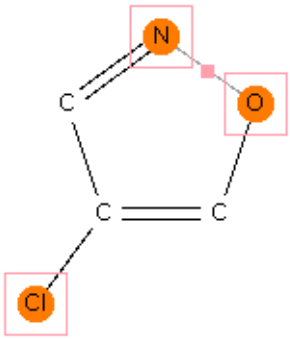
5.3.6 Undoing Mistakes when Drawing Substructures

- To undo the last action performed, click on **Edit** in the top-level menu and select **Undo** in the resulting pull-down menu.
- Alternatively, move the cursor to a blank point in the white area, click the right-hand mouse button, and select **Undo** from the pull-down menu.
- If necessary, **Edit**... **Undo** may be used several times in a row to undo a sequence of actions, one by one.

5.3.7 Selecting Atoms and Bonds

- Selection of atoms or bonds is useful for assigning properties such as element type and bond type, for moving substructures around the drawing area, and for cutting and pasting.

- A selected atom is coloured orange and enclosed in a box. If the two atoms at either end of a bond are selected then the bond itself is selected too. In the example below, the N, O and Cl atoms and the N-O bond are selected.



Atoms and bonds may be selected in several ways:

- In **EDIT** mode, an individual atom or bond can be selected by clicking on it with the left-hand mouse button.
- In **EDIT** mode, a series of atoms or bonds can be selected by clicking on each in turn while keeping the Shift key pressed down.
- In **EDIT** mode, a group of atoms and bonds can be selected by clicking with the left-hand mouse button on a blank point in the white area and moving the cursor while keeping the mouse button pressed down. Everything enclosed in the resulting rectangular box gets selected when the mouse button is released.
- In any mode, everything can be selected by hitting **Edit** in the top-level menu and **Select All** in the resulting pull-down menu.
- In any mode, the current selection can be reversed by hitting **Edit** in the top-level menu and **Invert Selection** in the resulting pull-down menu. Everything that was selected becomes unselected, and vice versa.

5.3.8 Deleting Atoms and Bonds

There are several ways of doing this:

- In **DRAW** mode, click with the right-hand mouse button on the atom or bond to be deleted and pick **Delete Atom** or **Delete Bond** from the resulting pull-down menu.
- Activate the **ERASE** mode (click on the **ERASE** button) and click with the left-hand mouse button on the atom or bond to be deleted.
- Activate the **EDIT** mode (click on the **EDIT** button). Select the atoms or bonds to be deleted (see Selecting Atoms and Bonds). Then either:
  - Click on **Edit** from the top-level menu and then **Cut** or **Delete Selected** from the resulting pull-down menu (**Cut** will delete both atoms and bonds, **Delete Selected** will give a choice).

- Click with the right-hand mouse button on a blank point in the white area and pick **Delete Selected** or **Cut** from the resulting pull-down menu.

- To delete all atoms and bonds, in **DRAW**, **EDIT** or **ERASE** modes, move the cursor onto a blank point in the white area, click on the right-hand mouse button, and pick **Clear All** from the pull-down menu.
- Use **Edit**... **Undo** (see Undoing Mistakes when Drawing Substructures) to delete an atom or bond just drawn.

5.4 Drawing and Fusing Rings

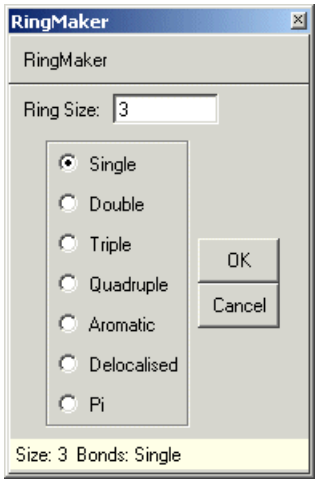
You can:

- Add a ring to a blank drawing area (see Adding a Ring to a Blank Drawing Area).
- Add a ring to an atom in an existing structure (see Adding a Ring to an Atom in an Existing Substructure).
- Fuse a new ring to an existing ring (see Fusing a New Ring to an Existing Ring).
- Create a spiro fusion (see Creating a Spiro-Fusion).
- Fuse two rings by moving one onto the other (see Fusing Rings by Moving One Ring onto Another).

5.4.1 Adding a Ring to a Blank Drawing Area

- Rings may be drawn manually but the easiest way is to use the pre-drawn rings in the bottom left-hand corner of the *Draw* window:

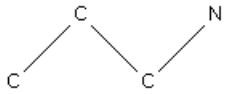
- If the desired ring is one of the four on display (see above), select it by clicking on the appropriate icon, move the cursor into the white area, then click with the left-hand mouse button.
- If you click again, you will create a second copy of the ring. Use **Edit... Undo** if this is not what was wanted. Click on **DRAW** to resume normal drawing.
- If a different ring size is required, hit **RingMaker**, type the desired ring size into the box, select the required bond type, hit **OK**, then click in the white area. In the example below, a saturated cyclopropane ring has been defined:



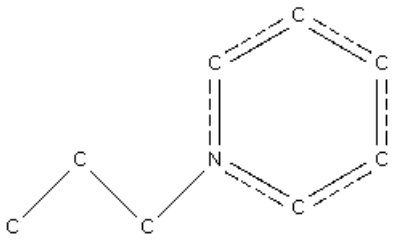
- Only carbon rings can be specified, but it is easy to change individual element types later (see Changing the Current Element Type).
- Some complex ring systems (for example crown ethers) are available by clicking on **Templates...** in the bottom left-hand corner.

5.4.2 Adding a Ring to an Atom in an Existing Substructure

- Select the ring (see Adding a Ring to a Blank Drawing Area), then click on the desired atom in the existing substructure with the left-hand mouse button.
- For example, selecting a 6-membered aromatic ring and clicking on the N atom in:

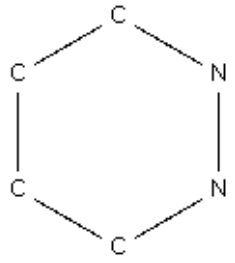


will create:

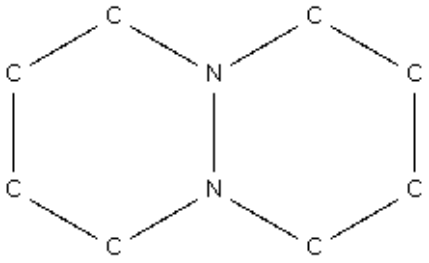


5.4.3 Fusing a New Ring to an Existing Ring

- Select the new ring (see Adding a Ring to a Blank Drawing Area), then click on the desired fusion bond in the existing ring.
- For example, selecting a 6-membered saturated ring and clicking on the N-N bond in:

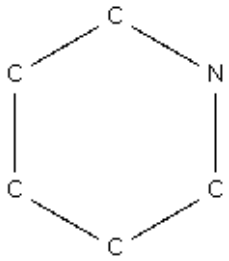


will create:

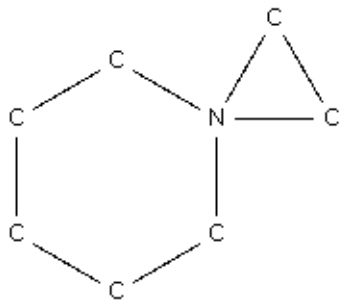


5.4.4 Creating a Spiro-Fusion

- Select the required ring (see Adding a Ring to a Blank Drawing Area), then click on the desired spiro atom in an existing ring.
- For example, selecting a 3-membered saturated ring and clicking on the N atom in:



will create:



5.4.5 Fusing Rings by Moving One Ring onto Another

It is possible to fuse two separate rings in the white area by selecting all the atoms in one ring (see Selecting Atoms and Bonds) and moving it towards the other (see Moving Atoms):

- Spiro fusion is achieved by overlapping one atom in the moveable ring with one atom in the stationary ring (indicated by the overlapped atoms going red). Fusion will occur when the mouse button is released.
- Bond fusion is achieved by overlapping two bonded atoms in the moveable ring with two bonded atoms in the stationary ring. It may be necessary to overlap one of the pairs and then rotate the moveable ring (by holding down the Control key) until the second pair overlap (see Rotating Structures).

5.5 Setting Atom Properties

You can:

- Change the current element type (i.e. the element type that will be assigned to any new atom created when drawing) (see Changing the Current Element Type)
- Change the element type of an existing atom (see Changing the Element Types of Existing Atoms)
- Automatically add hydrogen atoms (see Automatic Addition of Hydrogen Atoms)
- Add hydrogen atoms manually (see Adding Hydrogen Atoms Manually)
- Set atomic charges (see Setting Atomic Charges)

5.5.1 Changing the Current Element Type

- The current element type determines the type of any new atom created when drawing. It is displayed in the white box at the bottom of the *Draw* window. In the example below, the current element type is carbon:

- The current element type may be changed by hitting any of the element symbols at the bottom of the *Draw* window.
- Alternatively, it may be typed into the white box at the bottom of the *Draw* window.
- **More...** displays a pull-down menu. Selecting **Other Elements...** from this menu allows selection of any element in the Periodic Table.

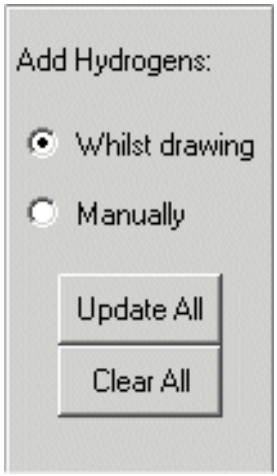
5.5.2 Changing the Element Types of Existing Atoms

This can be done in several ways, including:

- In **DRAW** or **EDIT** modes, click on the atom with the right-hand mouse button and select **Element** from the resulting pull-down menu. Then select the required element type. Selecting **More... then Other Elements...** allows selection of any element type in the Periodic Table.
- In **DRAW** mode, change the current element type (see Changing the Current Element Type) and then click on the atom with the left-hand mouse button.
- In **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, select **Element** from the resulting pull-down menu and select the required element (selecting **More...** then **Other Elements...** allows selection of any element type in the Periodic Table). The *Select Atoms* pop-up appears: click on the atom or atoms to be changed with the left-hand mouse button and hit **Done**.
- In **EDIT** mode, select the atom(s) to be changed (see Selecting Atoms and Bonds). Pick **Atoms** from the top-level menu, **Element** from the resulting pull-down menu, and select the required element. Selecting **More...** then **Other Elements...** allows selection of any element type in the Periodic Table.

5.5.3 Automatic Addition of Hydrogen Atoms

- All hydrogen atoms should be specified in order for Mogul to work reliably (see Required Molecular Information).
- Hydrogen atoms can be added automatically to structures. The number of hydrogens added to each atom will be sufficient to satisfy the atom's unfilled valencies.
- To automatically add hydrogen atoms whilst drawing a structure, either select the top-level menu button **Options** and ensure that **Auto-Generate H** is ticked, or select the **Whilst drawing** option in the *Add Hydrogens* area to the left of the *Draw* window:



- The appropriate number of hydrogen atoms will then be added or removed automatically whenever the element type of an existing atom is changed (see Changing the Element Types of Existing Atoms), a bond is drawn from an existing atom (see Drawing a Bond from an Existing Atom) or a bond is deleted (see Deleting Atoms and Bonds).
- To remove all hydrogen atoms, hit the **Clear All** button.
- To add hydrogen atoms automatically to all unfilled valencies in your structure, hit the **Update All** button.
- Occasionally, the program may assign the wrong number of hydrogens to an atom. This is especially likely for hydrogen-bridged metals, the oxygen atoms of metal-coordinated alcohols, and boron atoms in boron cages. It is advisable to check the number of H-atoms added to these types of structures. You can manually alter or explicitly define the number of hydrogens on an atom (see Adding Hydrogen Atoms Manually).

5.5.4 Adding Hydrogen Atoms Manually

- All hydrogen atoms should be specified in order for Mogul to work reliably (see Required Molecular Information).
- Select **Manually** in the *Add Hydrogens* area to the left of the *Draw* window to enable manual addition of hydrogens.
- Hydrogen atoms may be drawn in the same way as any other type of atom (see Fundamentals of Drawing) or they may be defined implicitly:
  - In **DRAW** or **EDIT** mode, click on an atom with the right-hand mouse button, pick **Hydrogens** from the resulting pull-down menu, then select the number of hydrogens required from the second pull-down menu. Alternatively, right-click on an atom, select **Hydrogens** from the resulting pull-down menu and **Generate** from the next menu.
  - In **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, hit **Hydrogens** in the resulting pull-down menu, then select the required number of hydrogens as above. The *Select Atoms*

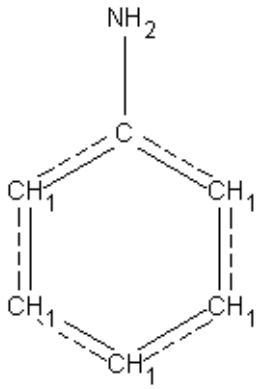
pop-up appears: click on the atom(s) to which hydrogens are to be added and hit **Done**. Alternatively, click on **Atoms** in the top-level menu, hit **Hydrogens** in the resulting pull-down menu, then hit **Generate** followed by **Selected Atoms**. Click on the atoms to which you want to add hydrogens and hit **Done**.

- In **EDIT** mode, select (see Selecting Atoms and Bonds) the atom(s) to which hydrogens are to be added. Then hit **Atoms** in the top-level menu, **Hydrogens** in the resulting pull-down menu, and then pick the required number of hydrogens. Alternatively, select the required atoms, click on **Atoms** in the top-level menu, hit **Hydrogens** in the resulting pull-down menu, then hit **Generate** followed by **Selected Atoms**.
- To add hydrogens to all atoms, go to **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, hit **Hydrogens** in the resulting pull-down menu, then hit **Generate** followed by **All Atoms.**

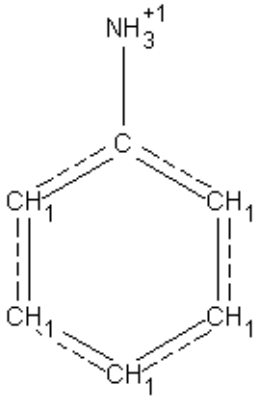
- To remove hydrogens, hit the **Clear All** button on the left of the *Draw* window. Alternatively, go to **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, hit **Hydrogens** in the resulting pull-down menu, then hit **Clear** followed by **All Atoms** or **Selected Atoms**. You can remove hydrogens from a particular atom by right-clicking on the atom and selecting **Hydrogens** and **Clear** from the resulting pull-down menus.

5.5.5 Setting Atomic Charges

- The formal (integer) atomic charge of any atom can be set in the *Draw* window.
- Atomic charges are ignored during a Mogul search. However, setting an atomic charge may affect the hydrogen count that is assigned by the program to that atom (see Automatic Addition of Hydrogen Atoms). For example, with the **Whilst drawing** check box selected in the *Add Hydrogens* pane, specifying an atomic charge of +1 on the N atom in:



will result in an anilinium ion with an increased number of hydrogen atoms on the N:



- The atomic charge of an atom can be specified in several ways:
  - In **DRAW** or **EDIT** mode, click on the atom with the right-hand mouse button, pick **Charge** from the resulting pull-down menu, then select the charge required from the resulting pull-down menus.
  - In **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, select **Charge** from the resulting pull-down menu, then pick the required charge as above. The *Select Atoms* pop-up appears: click on the atom(s) to which the charge is to be assigned and hit **Done**.
  - In **EDIT** mode, select the atom(s) to which a charge is to be assigned (see Selecting Atoms and Bonds). Then hit **Atoms** in the top-level menu, **Charge** in the resulting pull-down menu, and pick the required charge as above.

5.6 Setting Bond Types

You can:

- Change the current bond type (i.e. the type that will be assigned to any new bond created when drawing) (see Changing the Current Bond Type)
- Change the type of an existing bond (see Changing the Types of Existing Bonds)

5.6.1 Changing the Current Bond Type

- The current bond type determines the type of any new bond created when drawing (see Bond Type Conventions). The current setting is shown on the button next to the word *Bond* at the bottom of the *Draw* window.

- The current bond type may be changed by clicking on this button and selecting from the

resulting pull-down menu.

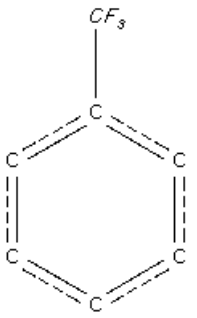
5.6.2 Changing the Types of Existing Bonds

This can be done in several ways, including:

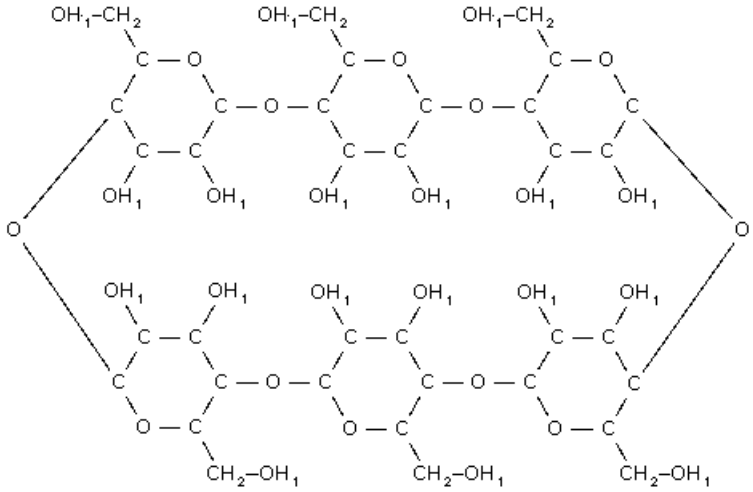
- In **DRAW** or **EDIT** modes, click on the centre of the bond with the right-hand mouse button and select **Type** from the resulting pull-down menu.
- In **DRAW** mode, change the current bond type (see Changing the Current Bond Type) and then click on the bond with the left-hand mouse button.
- In **DRAW** or **EDIT** mode, click on **Bonds** in the top-level menu, followed by **Type**, then select the required bond type from the resulting pull-down menu. The *Select Bonds* pop-up appears: click on the bond(s) to be changed with the left-hand mouse button and hit **Done**.
- In **EDIT** mode, select the bond(s) you want to change. Then pick **Bonds** from the top-level menu, followed by **Type**, and choose the required bond type from the resulting pull-down menu.

5.7 Using Chemical Groups and Structure Templates

Structure drawing can be made easier by using chemical groups and templates, both of which are pre-drawn structural fragments. Chemical groups are substituents such as -CF_3 which have a specific point at which they must be attached to the rest of the query structure, for example:

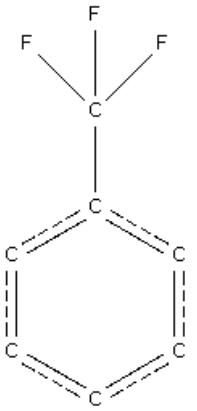


Templates are complete, stand-alone structural fragments, for example:



It is possible to:

- Add a chemical group to an existing structure (see Adding a Chemical Group to an Existing Atom).
- Replace a terminal atom in an existing structure by a chemical group (see Changing an Atom to a Chemical Group).
- Expand a chemical group so that all its atoms and bonds are shown in full (see Expanding a Chemical Group), for example:



- Access and add to the drawing area a standard templates, i.e. one of a set of templates supplied with Mogul (see Accessing Standard Structure Templates).
- Access and add to the drawing area a customised template, which you have drawn previously and saved for future use (see Saving and Using Customised Templates).

5.7.1 Adding a Chemical Group to an Existing Atom

This can be done in several ways, including:

- In **DRAW** or **EDIT** mode, right-click on the atom to which the group is to be added, select **Add Group**, then pick the group you want to add from the subsequent menus.
- In **DRAW** or **EDIT** mode, select **Atoms** from the top-level menu, then hit **Add Group** and pick the group you want from the subsequent menus. You will then be asked to pick the atom(s) to which the group is to be added.
- Select the atoms to which you want to add a group. Then hit **Atoms** in the top-level menu, **Add Group** from the resulting pull-down menu, and pick the group you require.
- In **DRAW** mode, change the current element type to a chemical group by hitting the **Groups...** button at the bottom of the window and selecting the required group from the subsequent menus. You can then draw groups in exactly the same way as you would normally draw an atom.
- A complete list of groups can be obtained by selecting the **View...** option, which appears in the menu that is displayed whenever an **Add Group** or **Groups...** button is hit. Groups may be selected from the resulting dialogue box and added to the query substructure.

5.7.2 Changing an Atom to a Chemical Group

This can be done in several ways, including:

- In **DRAW** or **EDIT** mode, right-click on an atom, select **Element** and then **Chemical Groups**, and then pick the desired group from the subsequent pull-down menus.
- In **DRAW** or **EDIT** mode, select **Atoms** from the top-level menu, then hit **Element** followed by **Chemical Groups**, and pick the group you want from the subsequent menus. You will then be asked to pick the atom(s) which are to be replaced by the group.
- Select the atoms which you want to replace. Then hit **Atoms** in the top-level menu, followed by **Element** and **Chemical Groups** in the resulting pull-down menus, and then pick the group you require.
- In **DRAW** mode, change the current element type to a chemical group by hitting the **Groups...** button at the bottom of the window and selecting the required group from the subsequent menus. Then left-click on the atom(s) you wish to replace.
- A complete list of groups can be obtained by selecting the **View...** option, which appears in the menu that is displayed whenever a **Chemical Groups** or **Groups...** button is hit. Groups may be selected from the resulting dialogue box and used to replace existing atoms.
- The program will not prevent you from changing a non-terminal atom to a chemical group, but it will never make chemical sense to do so, as a group only has one point of attachment (i.e. one unsatisfied valence).

5.7.3 Expanding a Chemical Group

- By default, chemical groups are shown as chemical symbols, for example *Et* for ethyl.
- A chemical group will be expanded automatically on selecting **Done** in the *Draw* window so that each atom and bond is shown explicitly within the *Build query* screen.

- To expand a chemical group within the *Draw* window right-click on the group and select **Expand Group**.
- Alternatively, select **Atoms** from the top-level menu, **Expand Chemical Groups** from the resulting pull-down menu, then either **Selected** or **All** (if you pick the former, you will be asked to pick the groups to be expanded).

5.7.4 Accessing Standard Structure Templates

- Several pre-drawn fragments are available to aid structure drawing.
- To view them, hit the **Templates...** button in the bottom left-hand corner of the *Draw* window, then select **View...** in the resulting pull-down menu. Alternatively, hit **File** in the top-level menu of the *Draw* window, followed by **Import Template** and then **View...**
- The resulting dialogue box not only shows the available templates but allows you to select a template and **Load** it into the drawing area.
- Alternatively, you can add a template to the drawing area by hitting the **Templates...** button in the bottom left-hand corner of the *Draw* window, selecting **List** in the resulting pull-down menu and then choosing the required structural type from the resulting pull-down menu (for example, **Boron-cages**). The exact template required can then be selected from the next pull-down menu (for example, **Hexaborane**).
- The same list of templates can also be accessed by hitting **File** in the top-level menu of the *Draw* window, followed by **Import Template** and then **List**.

5.7.5 Saving and Using Customised Templates

- To save the current contents of the white drawing area for future use as a template, pick **File** from the top-level menu followed by **Save Template...** in the resulting pull-down menu. The default file extension is *.cqt*.
- To read in a previously saved structural template, pick **File** from the top-level menu followed by **Import Template** and **File...** from the resulting pull-down menus.

5.8 Advanced Drawing Options

Advanced drawing options include:

- Moving atoms and structures (see Moving Atoms).
- Rotating structures (see Rotating Structures).
- Altering the size of structures (see Resizing Structures).
- Duplicating structures (copy, cut and paste) (see Duplicating Structures (Copy, Cut and Paste)).
- Changing default drawing options (see Changing Default Drawing Options).

5.8.1 Moving Atoms

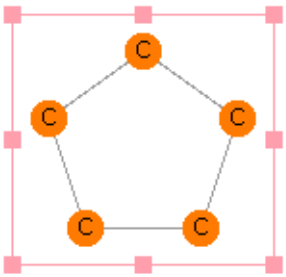
- In **EDIT** mode, select (see Selecting Atoms and Bonds) the atom(s) to be moved, move the cursor onto one of the selected atoms, press the left-hand mouse button, and move the cursor while keeping the button depressed.
- If two atoms overlap (coloured red) when the mouse button is released, they will be fused.

5.8.2 Rotating Structures

- It is only possible to rotate complete structures, not a collection of atoms which form part of a larger structure.
- In **EDIT** mode, select (see Selecting Atoms and Bonds) the structure to be rotated, move the cursor to a corner of the box surrounding the selected atoms, press the Control key and then the left-hand mouse button. The cursor should change shape to two curved arrows.
- Rotate by moving the cursor, keeping both the mouse button and the Control key depressed.

5.8.3 Resizing Structures

It is only possible to resize complete structures, not a collection of atoms which form part of a larger structure. In **EDIT** mode, select (see Selecting Atoms and Bonds) the structure to be resized, for example:



- To resize in the horizontal direction only, move the cursor to one of the solid pink squares at the middle of a vertical edge of the structure's bounding box, press the left-hand mouse button, and move the cursor left or right, keeping the mouse button depressed.
- To resize in the vertical direction only, move the cursor to one of the solid pink squares at the middle of a horizontal edge of the structure's bounding box, press the left-hand mouse button, and move the cursor up or down, keeping the mouse button depressed.
- To resize equally in both directions, move the cursor to one of the solid pink squares at a corner of the structure's bounding box, press the left-hand mouse button, and move the cursor, keeping the mouse button depressed.
- To resize unequally in both directions, move the cursor to one of the solid pink squares at a corner of the structure's bounding box, press the left-hand mouse button, and move the cursor,

keeping the mouse button and the Shift key depressed.

- If atoms overlap when the mouse button is released, they will not be merged.

5.8.4 Duplicating Structures (Copy, Cut and Paste)

To cause the program to take a copy of all or part of a structure, select (see Selecting Atoms and Bonds) the atoms and bonds to be copied and either:

- Click on a blank point in the white area with the right-hand mouse button and select **Copy** or **Cut** (**Cut** will delete the selected atoms and bonds, **Copy** will not); or
- hit **Edit** in the top-level menu and **Copy** or Cut in the resulting pull-down menu.

At this point, the copy of the selected atoms is held by the program but not placed into the drawing area. To do this, i.e. to paste the copy into the drawing area, either:

- Click on a blank point in the white area with the right-hand mouse button and select **Paste** (the option will be greyed out if there is nothing available to paste); or
- hit **Edit** in the top-level menu and **Paste** in the pull-down menu.

5.8.5 Changing Default Drawing Options

By default, all bonds drawn from an atom have a fixed length; other drawing options are also preset. These defaults can be changed using check buttons under the top-level menu item **Options.**

- To change the default length of bonds: ensure **Snap to Grid** is turned on and select **Drawing Options...** to change the value (the *Default bond length* is 60).
- To change the minimum increment that will be made to a bond angle as the cursor is moved when drawing a new bond from an existing atom: ensure **Snap to Grid** is turned on and select **Drawing Options...** to change the value (the *Default bond angle* is 15 degrees).
- To draw bonds freehand, i.e. to the exact position specified by the cursor movement, turn off the **Snap to Grid** check button.
- To change the tolerance for determining how close two atoms must be before they are judged to be overlapping, and how close the cursor must be before it is considered to be *on* an atom: select **Drawing Options...** and change the value (the default *Tolerance* is 15; increase to make the effective size of an atom bigger).

5.9 Pasting in Structures from ISIS/Draw and ChemDraw

Mogul offers the ability to copy and paste structures from ISIS/Draw (MDL Information Systems, Inc.) into the *Draw* window (Windows only). In order to use this facility, ISIS/Draw must be configured to copy a *MOL* file to the Windows clipboard:

- In ISIS/Draw, select **Options** from the top-level menu and **Settings...** from the resulting pull-down menu.
- Hit the *General* tab and switch on the **Copy Mol/Rxnfile to Clipboard** check-box.
- Hit **Save** if you wish this change to be applied to subsequent ISIS/Draw sessions.
- Hit **OK** to close the dialogue box and apply the change to your current ISIS/Draw session.

Now draw your structure in ISIS/Draw and copy all or part of the structure; select **Edit** from the ISIS/Draw top-level menu and **Copy** from the resulting pull-down menu.

To paste a copy of a structure drawn in ISIS/Draw into the *Draw* window, either:

- Click on a blank point in the white drawing area with the right-hand mouse button and select **Paste** (the option will be greyed out if there is nothing available to paste); or
- hit **Edit** in the top-level menu of the *Draw* window and **Paste** in the resulting pull-down menu.

It is also possible to copy and paste structures from ChemDraw (CambridgeSoft Corporation).

If Mogul is unable to translate atom or bond types specified in ISIS/Draw or ChemDraw, then these will be shown as *unknown* and may need to be edited in order to get the required search results.

5.10 Transferring the Current Structure into the *Build query* Pane and Closing the Draw Window

The *Draw* window can be closed down by:

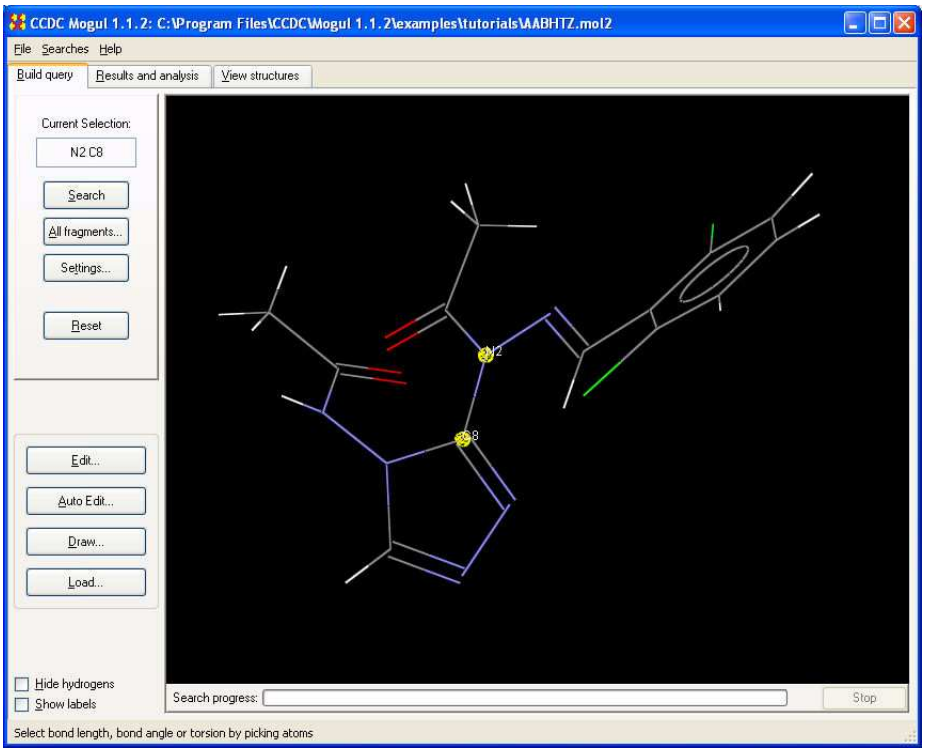
- Hitting the **Done** button (this transfers the current structure(s) to the *Build query* pane).
- Hitting the **Cancel** button or selecting **File** from the top-level menu and **Close** from the resulting pull-down menu. *Both of these options will discard the current contents of the drawing area.*

Note: once structures are transferred to the *Build query* pane, Mogul will standardise bond types and add any hydrogens it detects to be missing (see Assignment of Unknown Bond Types and Missing Hydrogen Atoms).

6 Running Searches

6.1 Searching for an Individual Bond Length, Valence Angle, Torsion Angle or Ring

- In the *Build query* pane, select the atoms that are needed to define the geometric parameter of interest by clicking on them with the left-hand mouse button (click on an atom again to deselect it).
- Select two atoms for a bond length (A-B), three consecutive atoms for a valence angle (A-B-C), four consecutive atoms for a torsion angle (A-B-C-D) or five or more atoms making up a ring. It is not possible to search for cyclic torsions, i.e. the central bond (B-C) of the torsion angle must be acyclic. It is not possible to search for bonds, angles or torsions in which one or more of A, B, C, D is a metal or hydrogen atom (however it is possible to search for rings containing a metal atom).
- It is possible to select rings for searching that are part of fused systems. However the ring libraries that are searched only contain examples of unfused rings. No fused ring systems containing the query will be returned.
- It is not possible to search for non-bonded contacts or improper torsions, so chemically-bonded atom sequences must be selected in order to form a valid search fragment.
- The current selection will be highlighted within the *Build query* display area. The atoms selected will also be listed under **Current Selection** on the left of the *Build query* pane:



- Hit **Reset** to clear the current selection.
- Hit **Search** to accept the current selection and run the search.

6.2 Searching for All Bond Lengths, Valence Angles, Torsion Angle and Rings

- An *All fragments* search will allow you to search for all valid bond lengths, valence angles, torsion angles and/or rings within your query (see Chemical Coverage: The Mogul Library).
- In order to perform an *All fragments* search your query should consist of complete molecules or ions only (see Query Structure Preparation).
- Hit **All fragments...** on the left of the *Build query* screen, and in the resulting *Search for all fragments* pop-up, select the geometric parameters that you wish to include. Select **Search** to run the search, or **Cancel** to return to the *Build query* screen.

6.3 Controlling the Number of Hits

6.3.1 Obtaining Additional Hits: Performing Generalised Searches

- When searching on a particular geometric feature, the number of hit fragments found that are structurally identical to the query fragment may not be enough. In such cases, Mogul can look for fragments that, while not identical to the query, are sufficiently closely related as to be relevant. These fragments may then be incorporated into a *generalised* distribution, which therefore consists of observations from fragments that are similar to the query fragment as well as fragments that are identical to the query (if these exist).
- Ring searches without generalisation, will only return rings that have the same distribution, size ansd stereochemistry of substituents as the query ring. The information used in regard to ring substituents is described earlier (see Section 2.2). If a ring search is generalised the atom types and order making up the ring are retained but the ring substituent sizes, number and setereochemistry can differ.
- To *prohibit* generalised searches, click on the **Settings...** button in the *Build query* pane. In the resulting *Mogul search settings* pop-up, select the *General tab* and switch on the **Find exact hits only** check-box. This will ensure that any hits found will be structurally identical to the query and will produce fast search speeds. However, it is still possible to find extra hits if there are not enough that are identical to the query fragment after the search has been run (see Finding More Hits).
- To *allow* generalised searches, ensure that the **Generalise** check-box is switched on. This may drastically lengthen search times - though this can be controlled to some extent (see Controlling the Speed and Quality of Generalised Searches) - but will greatly increase the chances of finding sufficient hits
  Note: Mogul will generalise searches by default if insufficient hits are found.
- The criteria used to control generalised searches can be set independently for bond, angle, torsion and ring searches. Select the appropriate tab in the *Mogul search settings* dialogue box. The following search settings are available:
  - **Generalise if less than** *X* **exact hits** is used to specify the minimum acceptable size of an exact

distribution (one containing only hit fragments that are structurally identical to the query). If the exact distribution does not contain at least this number of fragments, then Mogul will generalise the search, i.e. look for structurally related fragments. If **Generalise if less than** *X* **exact hits** is set to zero, Mogul will not generalise the search even if the exact distribution is empty.

- **Aim for at least** *X* **hits** is used to specify the number of observations that should be present in a generalised distribution. This number is a target which the search will aim for but may not exactly meet. If the number is set to zero, then Mogul will not perform a generalised search irrespective of how many hits it finds in the exact distribution.
- Hit fragments from a generalised search are ranked according to their *relevance* to the query fragment. Relevance values range from 0.0 to 1.0. If a fragment is identical to the query it will have a relevance of 1.0; otherwise it will have a relevance of less than 1.0 (the less relevant, the smaller the number). Only fragments with a relevance of at least the specified **Relevance threshold** value will be included in a distribution.
- It is recommended that a **Relevance threshold** of 0.75 or greater is used. If Mogul cannot find enough fragments satisfying the **Relevance threshold**, the size of the resulting distribution may be less than that requested
- When a generalised search is performed, the fragments that are included in the final distribution will depend on the *selection mode*. The choice of *selection mode* provides control over the size of the generalised distribution, the relevance of the fragments it contains, and the speed of the search (see Controlling the Speed and Quality of Generalised Searches).

To save your search settings between sessions, switch on the **Save settings on exit from Mogul** check-box under the *General tab* of the *Mogul search settings* dialogue. Individual search settings can be restored to their CCDC default values by clicking the **Default** button corresponding to that setting.

6.3.2 The Relevance Calculation and What it Encapsulates

- The relevance calculation is calculated in quite a complicated way. In summary the calculation first considers the core functionality of the feature in question and then looks at the chemistry further away from the feature. A feature which has essentially equivalent core structure will usually have a relevance of 0.8 or higher. The connectivity and atom types of the atoms directly connected to the geometric feature will also be matched to the query, and, if very similar, will raise the equivalence higher.
- For instance for the case of bonds, to be able to get a relevance of 0.8 or higher it is necessary that:
  - the atomic numbers, number of connections, highest bond order and hydrogen count of matching central atoms must be the same.
  - the bond order of the central bonds must be the same.
  - the size of any ring containing the bond must be the same.

- The final relevance value is determined by doing a weighted similarity calculation based on commonality between matched connected atoms. If there is a high degree of commonality between atomic numbers and highest bond orders for these then the relevance will tend to be higher. If atomic numbers match then this carries more weight than bond orders matching.
- A similar calculation is applied for bond angles and torsions.
- The relevance calculation for rings is slightly different. Ring matches of relevance 0.95 will have one atom of the ring differing from the query either in terms of number of substitutions (0, 1 or 2), substitution size (Small or large) or stereochemistry of substitution relative to other stereocentres in the molecule (up, down or in plane). Ring matches of lower relevance will have at least two atoms which differ from the query ring in these regards.

6.3.3 Controlling the Speed and Quality of Generalised Searches

- Exact searches are almost always very quick, but generalised searches may be much slower. The *selection mode* provides some control over the speed of generalised searches, the relevance of the hits they find, and the size of the resulting distribution.
- To specify selection mode, hit the **Settings...** button in the *Build query* pane. Selection mode can be set independently for bond, angle and torsion searches by selecting the appropriate tab in the resulting *Mogul search settings* dialogue box.
- As a *generalised* search progresses relevant fragments are identified. Either all of the hits found or just a subset of those hits can then be included in the final generalised distribution. The following options are available from the **Select subset** drop down menu:
  - **Optimise for relevance**: Mogul will try to find the most relevant fragments possible but may be slow. The size of the resulting distribution will usually be close to the value specified in the **Aim for at least** *X* **hits** entry box (see Obtaining Additional Hits: Performing Generalised Searches), though this is not guaranteed.
  - **Optimise for speed**: This gives the fastest generalised search speeds but will probably not find the most relevant fragments possible. However, all hit fragments will have a relevance at least as large as the specified **Relevance threshold** (see Obtaining Additional Hits: Performing Generalised Searches). Also, if there are any fragments in the Mogul library that are identical to the query (relevance = 1.0), they are guaranteed to be included in the distribution provided they satisfy other search criteria such as any specified filters. The size of the distribution will usually be close to the value specified in the **Aim for at least** *X* **hits** entry box (see Obtaining Additional Hits: Performing Generalised Searches), though this is not guaranteed. This is the default selection mode for generalised torsion searches.
  - **Speed/relevance compromise**: This is a compromise between **Optimise for relevance** and **Optimise for speed**. The distribution will be identical to that which would have been produced by **Optimise for relevance** unless this would necessitate Mogul looking at a large number of fragments, which would cause the search to be slow. If this happens, Mogul will give up trying to find the most relevant fragments and will accept any whose relevance is at least as

large as the specified **Relevance threshold** (see Obtaining Additional Hits: Performing Generalised Searches), provided they satisfy other search criteria such as any specified filters. This is the default selection mode for generalised bond, angle and ring searches.

- To include in the distribution all fragments whose relevance is at least as large as the specified **Relevance threshold** (see Obtaining Additional Hits: Performing Generalised Searches), provided they satisfy other search criteria such as any specified filters, select **Include all hits found**. This is the default setting. This may result in a large distribution containing a lot of fragments. The search may in some cases take a long time.

6.3.4 Imposing Level Limits on Generalised Searches

- When searching the knowledge base Mogul uses a set of structure keys that define a tree hierarchy. During a generalised search the algorithm will, by default, continue to proceed up the tree until the required number of hits are found that satisfy the relevance criteria. Therefore, occasionally Mogul can take a very long time to identify similar fragments when performing a generalised search. Limiting the number of levels traversed will reduce the chances of this happening but may also result in fewer hits being found.
- To impose a limit on the number of levels traversed, click on the **Settings...** button in the *Build query* pane. In the resulting *Mogul search settings* pop-up, select the *General tab* and switch on the **Impose upper level limits** button. Note: this option is only active if the **Generalised** radio button has been activated.
- Imposing upper level limits does not apply to ring searches. This is because a generalised ring search only goes one level up in the tree anyway.

6.3.5 Setting Search Filters

- Searches can be restricted e.g. to exclude structures with low experimental precision.
- Click on the **Settings...** button in the *Build query* pane, then in the resulting *Mogul search settings* pop-up, select the *Filters* tab. Available search filters include:
  - **R-factor:**
    Switch on the **R-factor** check box and then select one of the three options from the drop-down list in order to restrict subsequent searches to structures with R-factors less than or equal to 5%, 7.5% or 10%.
  - Exclusion of **Solvents** or **Non-solvents:**
    Switch on the **Exclude** check box and then select either **Solvents** or **Non-solvents** from the drop-down list. Subsequent searches will exclude fragments from either solvent or non-solvent molecules.
    Note: a molecule is considered to be a solvent if it is found in a CCDC catalog of known solvent molecules. However, if a crystal structure contains only solvent molecules then the largest molecule will be assumed to be a non-solvent.
  - **Heaviest Element:**

Switch on the **Heaviest Element** check box and then select an element from the drop-down list. Subsequent searches will exclude fragments from structures that contain elements heavier than the specified element.

- Exclusion of **Organometalics** or **Organics:**
  Switch on the **Exclude** check box and then select either **Organometalics** or **Organics** from the drop-down list. Selecting to exclude **Organometalics** will eliminate from the search any fragments from structures that contain a transition metal, lanthanide, actinide or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po. Selecting to exclude **Organics** will eliminate fragments from organic structures (i.e. structures that do not contain any of the elements listed above).

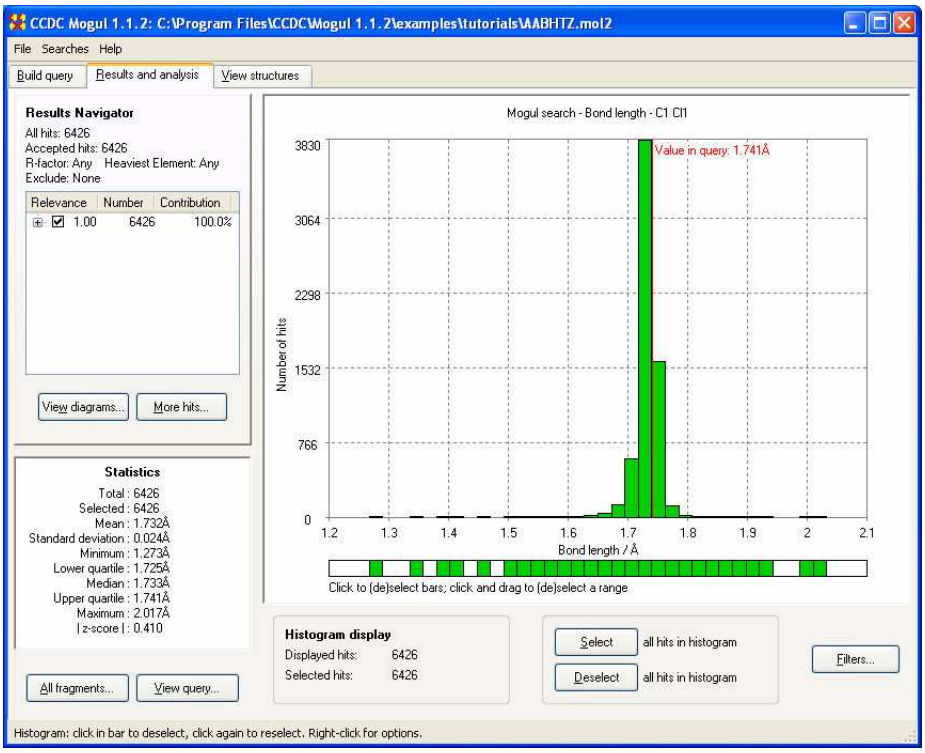
- To save your filter settings between sessions, switch on the **Save settings on exit from Mogul** check-box under the *General tab* of the *Mogul search settings* dialogue.
- Results can also be filtered after a search has been run (see Filtering Hits).

7 Viewing and Analysing Results

7.1 Viewing a Histogram and Selecting Hits

- On completion of a search, results are displayed as a histogram in the *Results and analysis* pane.
- For bond length, valence angle or torsion angle searches, the histogram shows the distribution of the respective geometric parameter calculated from the CSD entries that match your input query (i.e. those CSD entries that contain the same type of bond, angle or torsion as the query).
- The sign of a torsion angle calculated from a CSD entry is often arbitrary. For example, if the CSD entry is centrosymmetric, for every torsion angle with a positive sign there is, elsewhere in the unit cell, a symmetry-equivalent torsion with a negative sign. Consequently, only the absolute values of torsion angles are plotted in the histogram.
- For ring searches, the histogram shows the deviation of the ring conformation of each hit fragment from the query ring, in terms of the RMSD of the corresponding torsion angles. This RMSD is calculated as follows:
  - The search will only return rings of the same size and containing the same arrangement of Sybyl atom types as the query. Therefore it is possible to create a one-to-one mapping of the atoms of each hit ring to the query such that the Sybyl atom types match. A mapping between the bonds of the two rings is thus also implied.
  - For each bond of each ring, an "intracyclic" torsion can be calculated, i.e. the torsion angle formed by the bond in question and the two adjacent ring bonds.
  - For each mapped pair of bonds, the difference between the respective intracyclic torsions is calculated, taking into account the periodicity of the torsion measurement.
  - The RMSD (Root Mean Square Deviation) of the n torsion differences is then calculated.
  - If there is more than one way in which the Sybyl atom types can be matched (which is very often the case), then the RMSD is calculated for each possible mapping, and the smallest RMSD calculated is taken as the RMSD for the hit.
- **T**he absolute stereochemistry of a ring is often arbitrary, for the same reason as the sign of a torsion angle. Therefore, for each atom mapping, the RMSD calculation is performed twice, once with the hit ring in the conformation in which it is stored in the library, and once with it in an inverted conformation. Whichever result in smaller is taken as the RMSD for the mapping.
- *Note:* Unlike the bond, angle, and torsion libraries, where the geometric parameter is an inherent property of the fragment and is stored in the library, the ring RMSD is only meaningful in relation to a specific query ring. Following a ring search, the RMSD for each ring found in the search must therefore be calculated on-the-fly. This can sometimes take time for large subsets.
- *Note:* A consequence of using this RMSD measure is that an evaluation of the quality of the geometry in the query ring is accomplished by seeing how far the peak of the RMSD distribution is from 0 degrees. A peak position at 5 degress or less suggests good geometry of the query, a peak position of 10 degress or more may suggest an unusual ring geometry unless there are good examples with close to zero RMSD, which are very similar chemically speaking to the query.

- Distributions for very common types of fragments have been cut down by random selection to a subset of 10,000 observations. For rings this maximum distribution size is lower at 500 observations. This is because the similarity calculation made between the query ring and all the rings found, can take a long time for larger distributions
- If using a 3D input structure (as oppose to a 2D structure), the value of the bond length, valence angle or torsion angle in the query is superimposed in red on the histogram display to allow for easy comparison with the geometric results obtained from Mogul:

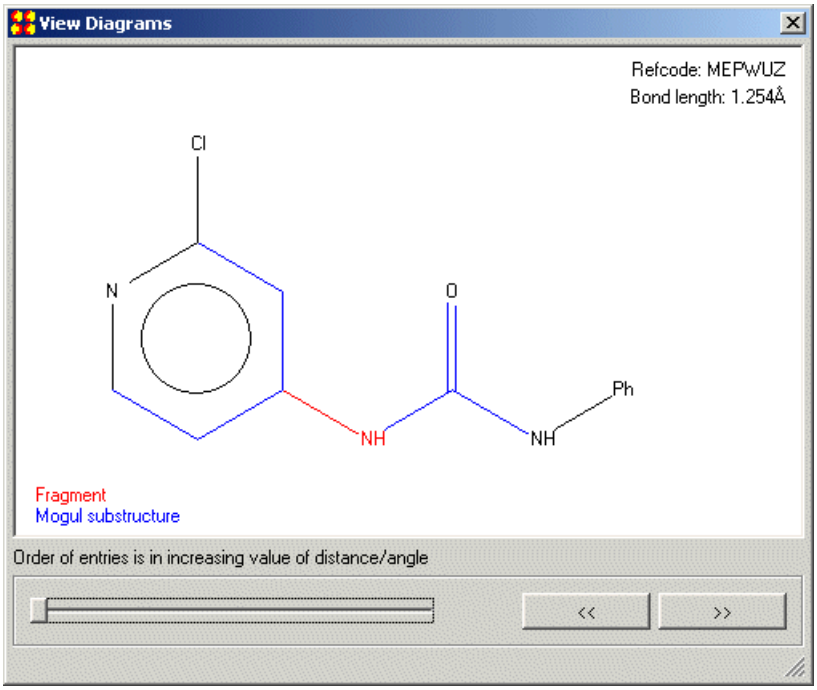


- By default, all the CSD entries contributing to a histogram are listed and can be viewed in the *View structures* pane. However, this list can be restricted to the CSD entries that contribute to one or more chosen bins by selecting those bins in the histogram.
- To select or deselect individual bins, click on them with the left mouse button (each click on a bin will reverse its selection status).
- Histogram bins can also be selected or deselected by using the horizontal bar located directly under the histogram. Click on the element of the bar directly under the required bin in order to reverse that bin's selection status. To reverse the selection status of a range of bins, position the cursor on the element of the horizontal bar directly underneath the first of the required bins, press down the left-hand mouse button and *while keeping the mouse button depressed* move the cursor over the required range.
- Summary statistics for selected hits in the histogram are displayed on the left hand-side of the *Results and analysis* pane. These statistics are automatically updated to reflect the hits currently selected (see Summary Statistics).

- A *Histogram display* is shown under the histogram and will also be updated as bins are selected and deselected.
- To change the display style of a histogram right-click in the histogram display area, and from the resulting menu select the required option (**Font..., Selected Colour..., Deselected Colour..., Background Colour...**).

7.2 Browsing the Chemical Structures of Hits

- Select the **View diagrams...** button to the left of the *Results and analysis* pane to browse the chemical structures of those hits currently selected within the *Results Navigator* (see Viewing and Navigating Generalised Search Results).
- The resulting *View Diagrams* pop-up allows the 2D chemical diagram of each hit structure to be viewed. The fragment (bond, angle or torsion) in the hit structure that matches the query fragment is, by default, highlighted in red. The other parts of the hit structure that were taken into account in the search are, by default, highlighted in blue. To change the display style, right-click and select the appropriate option from the pull-down menu.
- The CSD refcode (entry ID) and the value of the geometric parameter of interest are also displayed:



- The individual hits are displayed in order of increasing value of bond length, valence angle, torsion, or, foir rings, torsion RMSD. The buttons labelled << and >> can be used to step through the structures, backwards or forwards, one by one.
- Alternatively, the *slider bar* can be used to scan through structures more rapidly. Either press down the left-hand mouse button on the slider handle and *while keeping the mouse button*

depressed move the handle along the bar, or press down the left-hand mouse button on the slider bar itself *and keep the mouse button depressed*.

7.3 Filtering Hits

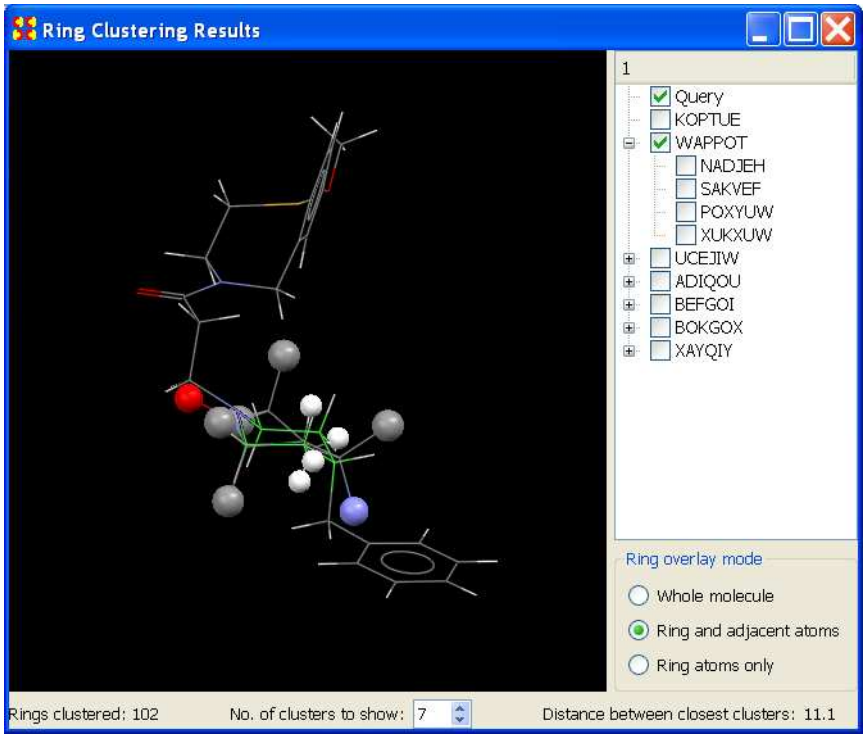
- Search results can be filtered e.g. to exclude structures with low experimental precision.
- Click on the **Filter** button at the bottom of the *Results and analysis* pane.
- In the resulting *Apply Filters* pop-up select the filters you wish to use. Available filters include:
  - **R-factor:**
    Switch on the **R-factor** check box and then select one of the three options from the drop-down list in order to restrict the results to structures with R-factors less than or equal to 5%, 7.5% or 10%.
  - Exclusion of **Solvents** or **Non-solvents:**
    Switch on the **Exclude** check box and then select either **Solvents** or **Non-solvents** from the drop-down list. Results will be filtered to exclude fragments from either solvent or non-solvent molecules.
    Note: a molecule is considered to be a solvent if it is found in a CCDC catalog of known solvent molecules. However, if a crystal structure contains only solvent molecules then the largest molecule will be assumed to be a non-solvent.
  - **Heaviest Element:**
    Switch on the **Heaviest Element** check box and then select an element from the drop-down list. Results will be filtered to exclude fragments from structures that contain elements heavier than the specified element.
  - Exclusion of **Organometalics** or **Organics:**
    Switch on the **Exclude** check box and then select either **Organometalics** or **Organics** from the drop-down list. Selecting to exclude **Organometalics** will eliminate from the results any fragments from structures that contain a transition metal, lanthanide, actinide or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po. Selecting to exclude **Organics** will eliminate fragments from organic structures (i.e. structures that do not contain any of the elements listed above).
- The selected filter(s) can be applied either to the displayed results only or to all results from the query molecule (i.e. when using an *All fragments* search).
- Hit **OK** to accept the current filters. Search results will update immediately to reflect the chosen filter(s). Alternatively, hit **Cancel** to return to the *Results and analysis* pane.

Note: If any filters were set before running a particular search (see Setting Search Filters) then it will not be possible to retrospectively display excluded entries.

7.4 Clustering the Results of Ring Searches

- If a ring search has been carried out the **Cluster** button at the bottom right corner of the *Results and analysis* pane becomes active.

- Clicking on this button brings up a window in which a cluster analysis of the ring search can be carried out. *Note:* If the number of hits to be clustered is greater than a hundred, you will be asked whether you wish to reduce this number. This is because the ring clustering algorithm calculates a 'distance' between each pair of hits and so a large sets of hits take an excessively long time to cluster.
- The structure of the query compound will be visible. This can be toggled on and off as required.
- Initially only a single ring structure from the hit list is shown. This is the most representative entry in the CSD for the full set of search results.
- To increase the number of clusters from 1, use the arrows at the bottom middle of the window. The maximum number of clusters is equal to the number of hits in the result set.



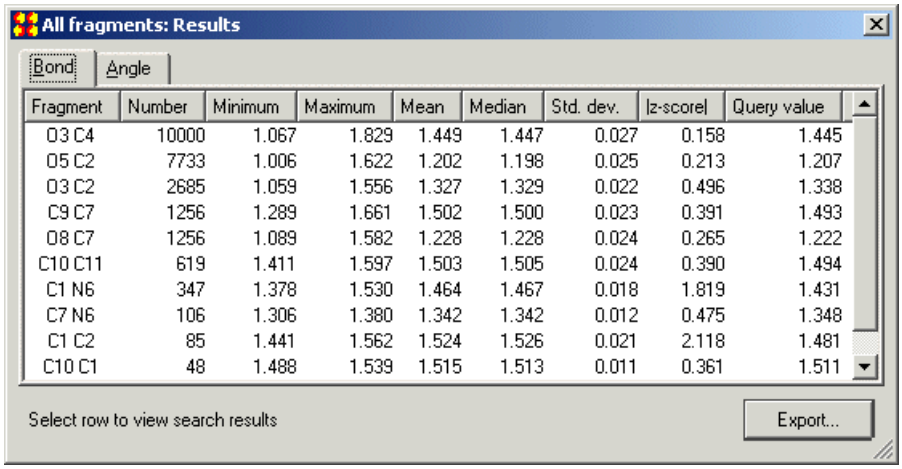
- The clustering algorithm uses a distance metric between each ring pair. The distance between each pair of rings is the RMSD of the torsion angles, calculated in the same way as the values that are plotted on the histogram.
- The distances are then used as the input to a standard complete linkage agglomerative clustering algorithm. (i.e. Every ring is first considered as a singleton, then the two most similar rings (i.e. smallest RMSD) are merged into a single cluster. Then the two closest clusters, are merged where the distance between two clusters is defined as the largest pairwise RMSD of any pair of individual rings taken from the two clusters. This is repeated until the number of clusters required is attained).
- The distance between closest clusters is displayed at the bottom-right of the window.
- The most representative structure of each cluster is superimposed. Each cluster can be expanded

via the tree expansion + boxes in the pane at the top-right. Additional cluster memberes can then also be displayed.

- The overlays can be displayed in three ways using the toggle buttons at the bottom-right. Full molecule overlays can be viewed, or the view can be restricted to the rings plus adjacent atoms, or only the ring atoms themslves can be displayed.

7.5 All-Fragments Search Results

- An *All-fragments* search will return results for each bond, valence angle, torsion and non-fused ring within your query, apart from those that are not represented in the Mogul library (see Chemical Coverage: The Mogul Library)).
- The results from an *All-fragments* search are displayed in a separate window. The *All-fragments search* window will open automatically upon completion of an *all-fragments* search. Once closed, this window can be re-opened by clicking on the **All fragments...** button to the left of the *Results and analysis* pane.
- To view results for either bonds, valence angles, torsions or rings, click on the appropriate tab at the top of the *All-fragments search* window:



- Results are displayed in a spreadsheet. Each search fragment is listed in the spreadsheet together with its associated summary statistics (see Summary Statistics).
- The rows of a spreadsheet can be sorted according to the values in any of the columns (see Manipulating Spreadsheets).
- To export results for the current fragment type (i.e. bond, angle or torsion) as a text file, click on the **Export...** button in the *All-fragments search* window (see Exporting All-Fragments Search Results).
- Full details of the results for a particular fragment can be displayed in the *Results and analysis* and *View structures* panes of the main *Mogul* window by clicking on the corresponding row in

the spreadsheet. The chosen fragment will be highlighted in the *Build query* structure display.

7.5.1 Summary Statistics

Summary statistics are listed in the *Statistics* box in the bottom left-hand corner of the *Results and analysis* pane. The statistics shown will depend on whether you are viewing results for a bond length, valence angle, torsion angle or ring search.

Statistics for bond and angle search results include:

- Number of observations, minimum, maximum, mean, median and standard deviation.
- z-score, which is the absolute difference between observed and mean values of a geometric parameter divided by the standard deviation of the Mogul distribution. A high value (e.g. >2.0) may indicate unusual or even suspect geometric features within your query.

Statistics for torsion angle search results include:

- Number of observations, minimum and maximum.
- d(min), which is the difference between the value of the torsion angle in your query and the nearest torsion angle to it in the Mogul distribution.

Note: z-score and d(min) are only available when using a 3D input query.

Statistics for rings include only:

- Number of observations, minimum and maximum torsion RMSD.

7.5.2 Manipulating Spreadsheets

- The rows of a spreadsheet may be sorted according to the values in any column by clicking on the column header. For example, to sort the rows of a *Bond* spreadsheet by z-score, click on the **z-score** button at the top of the spreadsheet. To reverse the order, click on the **z-score** button again.
- The order in which spreadsheet columns are displayed can be changed by pressing down the left-hand mouse button on a column header, moving the cursor *while keeping the mouse button depressed* to the new location, and then releasing the button.

7.6 Finding More Hits

- The number of hit fragments found that are structurally identical to the query fragment may not be enough. In such cases, Mogul can look for fragments that, while not identical to the query, are sufficiently closely related as to be relevant. These fragments may then be incorporated into a *generalised* distribution, which therefore consists of observations from fragments that are similar to the query fragment as well as fragments that are identical to the query (if these exist).
- To find additional hits (i.e. perform a *generalised* search) for the current search fragment click on the **More hits...** button in the *Results and analysis pane*. The resulting *Mogul: Find more hits*

window contains a number of settings that are used to control the *generalised* search. Mogul will identify related fragments to include in the *generalised* distribution using the following criteria:

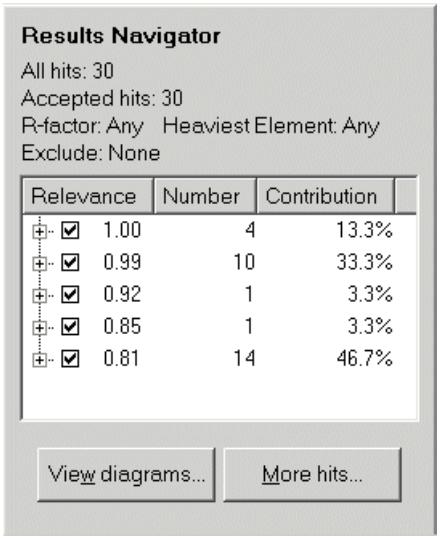
- **Aim for at least** *X* **hits** is used to specify the number of observations that should be present in a generalised distribution. This number is a target which the search will aim for but may not exactly meet. If the number is set to zero, then Mogul will not perform a generalised search irrespective of how many hits it finds in the exact distribution (i.e. without generalisation).
- When a generalised search is performed, the fragments that are included in the final distribution will depend on the **Selection mode**. The choice of selection mode provides control over the size of the generalised distribution, the relevance of the fragments it contains, and the speed of the search (see Controlling the Speed and Quality of Generalised Searches).
- Hit fragments from a generalised search are ranked according to their *relevance* to the query fragment. Relevance values range from 0.0 to 1.0. If a fragment is identical to the query it will have a relevance of 1.0; otherwise it will have a relevance of less than 1.0 (the less relevant, the smaller the number). Only fragments with a relevance of at least the specified **Relevance threshold** value will be included in a distribution. It is recommended that a **Relevance threshold** of 0.75 or greater is used. If Mogul cannot find enough fragments satisfying the **Relevance threshold**, the size of the resulting distribution may be less than that requested.
- **Search Filters** can be used to exclude e.g. fragments from structures with low experimental precision from the *generalised* search. A number of different filters are available (see Section 6.3.5).
  Note: Any filter(s) specified for the *generalised* search will also be applied to the current *exact* search results and it will not be possible to retrospectively display excluded entries.

- Select **OK** to run the search using the current criteria or **Cancel** to return to the *Results and analysis* pane.
- It is also possible to configure search settings to enable generalised searching in advance of running a search (see Controlling the Number of Hits).

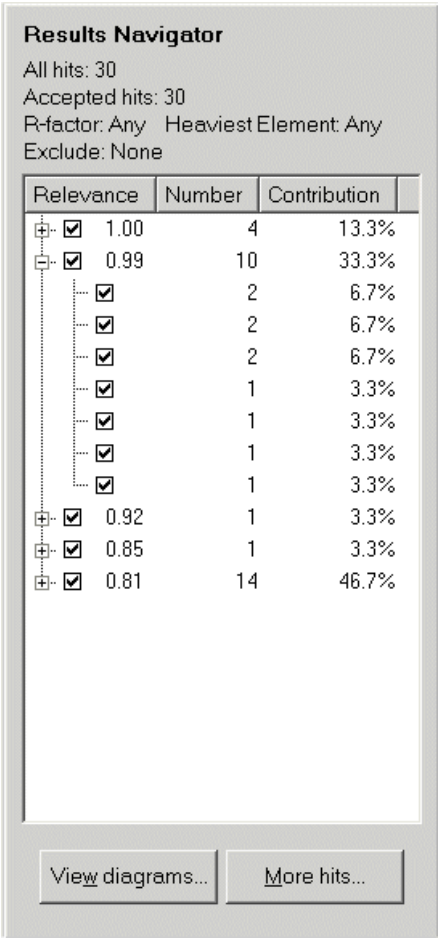
7.7 Viewing and Navigating Generalised Search Results

- When searching on a particular geometric feature, the number of hit fragments found that are structurally identical to the query fragment may not be enough. In such cases, Mogul can look for fragments that, while not identical to the query, are sufficiently closely related as to be relevant (see Obtaining Additional Hits: Performing Generalised Searches). These fragments may then be incorporated into a final, *generalised* distribution.
- A generalised distribution therefore consists of observations from fragments that are similar to the query fragment as well as fragments that are identical to the query fragment (if these exist). Mogul determines how closely related a particular fragment is to the query fragment by calculating its "relevance" (a number between 0 and 1; higher values imply closer similarity to the query).
- Fragments contributing to a generalised distribution are listed in the **Results Navigator** located

on the left of the *Results and analysis* pane. By default, fragments are listed in descending order of relevance (so any exact matches will come first, with a relevance of 1.0). The number of fragments with a given relevance is also shown, together with their percentage contribution to the total distribution:



- The information in the **Results Navigator** can be sorted according to the values in each of the columns by clicking on the **Relevance**, **Number** or **Contribution** column header button. To reverse the order, click on the column header button again.
- Groups of contributing entries can be switched on and off using their corresponding check-box within the **Results Navigator**. The histogram and summary statistics are automatically updated to reflect the fragments currently selected.
- It is possible for structurally different fragments to have the same relevance value. Contributions from such fragments can be viewed and individually selected by clicking on the corresponding expansion icon:



- The chemical structures of those fragments currently selected can be viewed by clicking on the **View diagrams...** button (see Browsing the Chemical Structures of Hits).
- In order to modify the search settings used (e.g. to find additional hits) click on the **More hits...** button (see Finding More Hits).

7.8 Viewing the Query Structure and Search Settings

- Select the **View query...** button in the *Results and analysis* pane to display information about the query structure and search settings in a separate window.
- In the resulting *Mogul: Query* window, select the *Molecule* tab to display the query structure. Detaching the query in this way enables the query structure and either the *Results and analysis* or *View structures* pane to be viewed simultaneously. The search fragment (bond, angle, torsion or) will be highlighted in the structure. Right-clicking in the visualiser display area will provide access to many options (available for 3D query structures only) (see Using the 3D Visualiser).
- Select the *Settings* tab in the *Mogul: Query* window to display information on the current search settings including:
  - Information on the molecule input file and search fragment.

- The criteria used to control the number of hits retrieved, the speed of the search and the relevance of the hits found (see Controlling the Number of Hits).
- Information on any filters applied to the search itself and the subsequent display of results

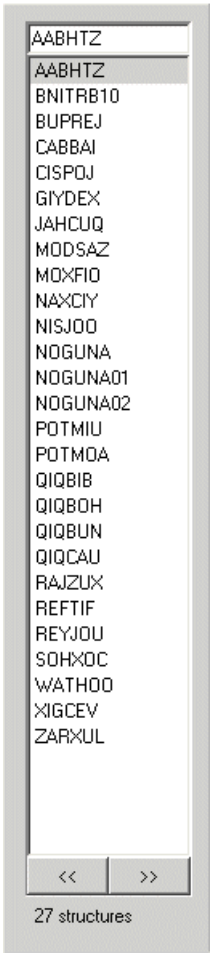
7.9 Viewing Searches Done Earlier in a Session

- If you have done two or more searches in a Mogul session, whether on the same query structure or not, you can go back to one of the earlier searches by selecting the top-level menu item **Searches** and then selecting the search you want from the resulting menu.
- If you have run an all-fragments search (see Searching for All Bond Lengths, Valence Angles, Torsion Angle and Rings) on the current query structure but have closed down the spreadsheet of results, you can get it back by clicking on the **All-fragments results...** button in the *Results and analysis* pane.

8 Viewing Hit Structures

8.1 Selecting Individual Structures for Viewing

- Any structure found by a Mogul search can be selected for viewing by clicking on its refcode (CSD entry identifier) in the list on the right-hand side of *View structures*.
- Only one structure can be viewed at a time and the structure currently on display is highlighted:

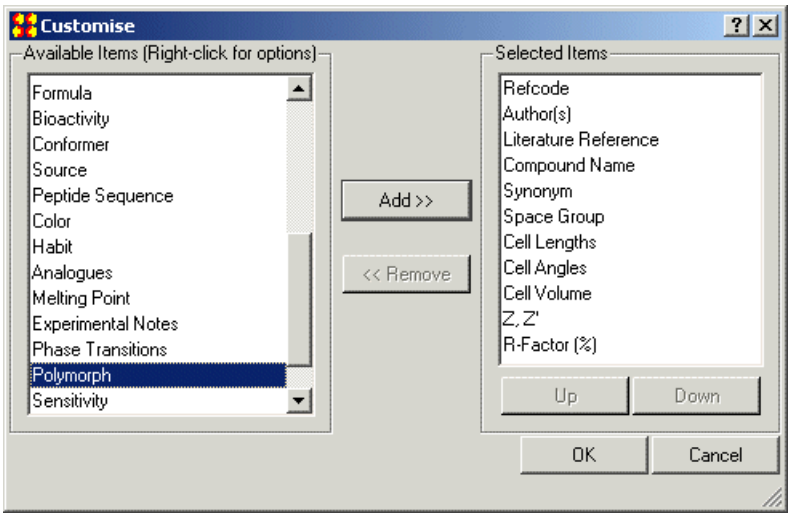


- All selected structures - i.e. structures corresponding to currently selected bars of the histogram (see Viewing a Histogram and Selecting Hits) - are included in the refcode list. The total number of these structures is displayed under the list.
- The buttons labelled << and >> can be used to step through the structures, backwards or forwards, one by one.

8.2 Displaying Information about Individual Structures

- Information such as the literature reference, chemical name, etc. for the currently selected structure (i.e. the one highlighted in the *View structures* hit list) may be available; this information can be viewed by hitting the **Information** button in the *View structures* screen.

- The specific data items displayed can be changed by clicking on the **Customise** button in the bottom-left corner of the screen.
- This opens a window showing two lists. The right-hand list shows those data items that are currently being displayed. The left-hand list shows other data items that it is possible to display. If an item occurs in the right-hand list, it will be absent from the left-hand list, and vice versa.

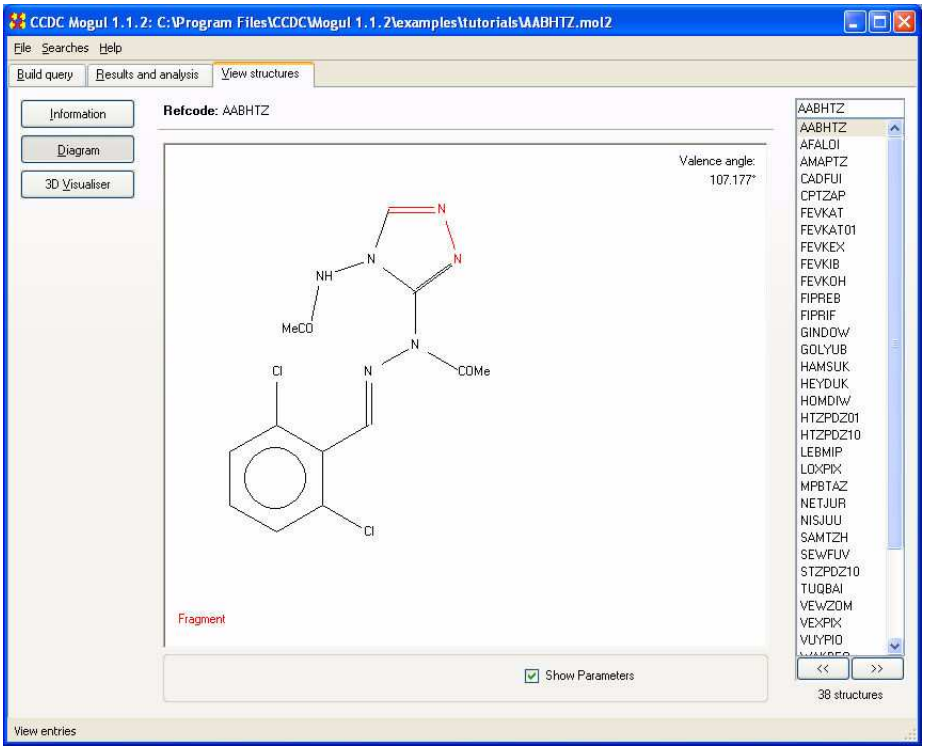


- To add a new item, select the required item by clicking on it in the left-hand list, then hit the **Add** button.
- To remove an item, click on it in the right-hand list and hit the **Remove** button.
- Alternatively, double-click on an item to transfer it from one list to the other.
- To change the order in which the items are arranged, click on an item in the right-hand list and use the **Up** or **Down** button to alter its position in the list.
- Using the **Add**, **Remove**, **Up** and **Down** buttons will not change the corresponding items displayed in the *View Structures* pane until the **OK** button is hit to close the *Customise* dialogue window.

Note: A list of the available data items is given in **Appendix C: Information Available for Individual Structures**.

8.3Viewing the 2D Chemical Diagram

- Each hit structure has a 2D chemical diagram; this can be seen by clicking on the **Diagram** button in the *View structures* screen:



- The Mogul search fragment (i.e. the atoms in the hit structure that match the atoms in the query used to define the geometric parameter of interest) will be highlighted in red.
- Switch on the **Show Parameters** check box to display the value of the geometric parameter corresponding to the highlighted fragment. This check-box may be found directly under the chemical diagram display area. The geometric value is displayed in the top right corner of the chemical diagram area.
- If the search fragment occurs more than once in the structure, then the number of matching hits will be displayed directly under the chemical diagram display area. Click on the up and down arrows alongside the hit number to highlight each matching search fragment in turn.

8.4 Using the 3D Visualiser

8.4.1 Visualiser Basics

- Hit *3D Visualiser* in *View structures* to see the currently selected structure in 3D.
- Use the mouse buttons to move the molecule (see Rotating, Translating and Scaling in the 3D Visualiser).
- Right-click anywhere in the visualiser area to produce the menu for measuring geometry, changing display styles and generating packing diagrams (see Right-Clicking in the Visualiser Display Area).
- The display area can be returned to the default view (viz. looking down the crystallographic *b*

axis), scale and perspective by clicking on the **Reset View** button under the 3D display area.

8.4.2 Right-Clicking in the Visualiser Display Area

Right-clicking in the 3D display area generates menus that provide access to many options. The menu you see will depend on whether you right-click on:

- A blank area in the display window, away from objects such as atoms, or bonds.
- Specific objects in the display.

Menu items will be greyed out if they are inapplicable given the current state of the display.

8.4.3 Selecting and Deselecting Atoms and Molecules in the 3D Visualiser

Selection of atoms and molecules is useful for changing properties such as display style. Atoms may be selected or deselected in several ways:

- Right-click anywhere in the 3D display area (atom, bond or background) and choose **Selection** from the resulting menu. Various options for selecting atoms can then be chosen from the resulting menus.
- Right-click in the display-area background, select **Picking Mode** from the resulting menu and set the *Picking Mode* to **Default Picking Mode**. Click on individual atoms with the left mouse button to select them. Once selected, an atom can be deselected by clicking on it again.
- When the *Picking Mode* is set to **Default Picking Mode**, all atoms become deselected if you left-click anywhere in the display-area background.
- Press the *shift* key and then click on any atom to select or deselect the entire molecule containing that atom.

8.4.4 Rotating, Translating and Scaling in the 3D Visualiser

- **Movement:** Rotate by moving the cursor around in the 3D window while keeping the left-hand mouse button pressed down. Rotate around the z-axis (the axis perpendicular to the screen) by keeping the left-hand mouse button *and* the Shift key pressed down. Translate by moving the cursor in the 3D window with the centre mouse button depressed (requires 3-button mouse). Alternatively, use the left-hand button with the Control key pressed down.
- **Scale**: Zoom in or out by moving the cursor up and down in the 3D window while keeping the right-hand mouse button pressed down.
- **Rotation Centre:** The centre of rotation can be set to a specific point or atom by right-clicking on that point or atom, selecting **Rotation Centre** from the resulting menu, and then choosing **Set Rotation Centre**. Select **Reset Rotation Centre** to revert to the default centre of rotation.

8.4.5 Viewing Along Particular Directions in the 3D Visualiser

- **Viewing Along Crystallographic Axes:** The contents of the display area can be viewed along real cell axes or reciprocal cell axes by right-clicking in the background, selecting **View** from the

resulting menu, and then choosing the view direction from the resulting menu.

- **Viewing Along or Perpendicular to a Bond:** This can be done by right-clicking on the bond and selecting **View along bond** or **View perpendicular to bond**.

8.4.6 Changing Display Styles in the 3D Visualiser

- A choice of four display styles are offered: wireframe, capped stick, ball-and-stick, and space-filling.
- To change the display style of the whole structure, right click anywhere in the visualiser box background. Select **Styles** from the resulting pull-down menu and then pick the required style.
- To change the style of just a few atoms, select them (see Selecting and Deselecting Atoms and Molecules in the 3D Visualiser), right-click anywhere to get the pull-down menu, hit **Styles** and pick the required style.
- Drag the right-hand mouse button while keeping the Shift key depressed to change the *field of view*, i.e. to go from orthographic projection (the default) to perspective projection with increasingly large viewing angles.
- To identify aromatic rings by displaying a circle within them, right-click anywhere in the visualiser background and select **Display Aromatic Rings** from the resulting menu.

8.4.7 Setting a Global Colouring Scheme in the 3D Visualiser

- Atoms can be coloured by element type or symmetry equivalence; right-click in the display-area background, pick **Colours** from the pull-down menu, and select the required option (**Custom Carbon**, **Colour By Element, Colour by Symmetry Equivalence, Colour by Atomic Displacement).**
- **Custom Carbon** alows a colur other than the default grey to be chosen for the carbon atoms within the molcule
- When atoms are coloured by element, each atom is assigned a colour depending on its element type.
- If all atoms are coloured by symmetry equivalence, each different molecule (or ion) in the crystal chemical unit is assigned a different colour (the crystal chemical unit is the same as the asymmetric unit in the majority of structures).
- If all atoms are coloured by atomic displacement then the thermal motion of the atom will be designated by the colour via a spectrum dsiplay (blue: low, red: high). CSD entries do not have atomic displacements stored currently.
- When a packing diagram is then constructed, any given molecule is assigned the same colour as that of the molecule in the crystal chemical unit to which it is related by crystallographic symmetry.
- This means that molecules of the same colour are crystallographically (and therefore chemically and geometrically) identical to each other.

- You can switch between the default black background and an alternative colour by right-clicking in the background area and hitting **Draw Backdrop**. The alternative colour will be a blue gradient.

8.4.8 Setting Display Properties for Particular Atom(s) or Bond(s) in the 3D Visualiser

To set the display properties of a particular atom, bond, molecule, or set of atoms:

- Select the atoms whose display properties you wish to change (see Selecting and Deselecting Atoms and Molecules in the 3D Visualiser). Right-click in the display-area background, select the appropriate option from the pull-down menu (**Styles, Colours, Labels, Show/Hide**) and then select the desired display-property setting from the next menu.
- Alternatively, right-click on an individual atom or bond, select the appropriate option from the pull-down menu (**Styles, Colours, Labels, Show/Hide**) and then select the desired display-property setting from the next menu. The chosen setting will be applied to the atom or bond on which you clicked

8.4.9 Labelling Atoms in the 3D Visualiser

- Right-click anywhere in the visualiser background and select **Labels** from the resulting pull-down menu, then hit **Show Labels** to label the atoms currently selected (see Selecting and Deselecting Atoms and Molecules in the 3D Visualiser). If no atoms are selected, then all atoms in the structure will be labelled.
- To remove all labels, right-click in the visualiser background, select **Labels** from the resulting pull-down menu, then hit **Hide Labels.**
- You can switch individual atom labels on and off by right-clicking in the display-area background and setting *Picking Mode* to **Pick Labels** in the resulting menus. When the **Pick Labels** option is active, left-clicking on an atom will then toggle its label on and off.

8.4.10 Measuring Distances, Angles and Torsions

To measure distances, angles and torsions:

- Right-click anywhere in the visualiser background, select either **Measure** (or **Picking Mode)** from the resulting pull-down menu, then select either **Measure Distances, Measure Angles** or **Measure Torsions** (or **Pick Distances, Pick Angles, Pick Torsions**). Depending on which mode has been chosen, you can then click on two, three or four atoms, respectively, to measure a distance, angle or torsion.
- You will remain in the chosen measurement mode, so after measuring the first distance (or angle or torsion), you can continue measuring others.
- To cancel the measurement mode, right-click anywhere in the background and select **Measure** or **Picking Mode** followed by **Default Picking Mode**.
- To clear all measurements from the display, right-click anywhere in the background and select **Clear Measurements**.

8.4.11 Displaying Crystallographic Unit-Cell Contents

- Right-click anywhere in the visualiser background, select **Packing** from the resulting pull-down menu and then **Packing** from the menu that follows.
- To go back to displaying the molecule, right-click in the visualiser background and select **Packing** followed by **Molecule.**

8.4.12 Displaying Search Fragments in the 3D Visualiser

- Switch on the **Display fragments** check box to highlight the Mogul search fragment (i.e. the atoms in the structure that match those atoms in the query used to define the geometric parameter of interest). The value of the associated geometric parameter will also be displayed. This check-box may be found directly under the 3D visualiser area.
- If the fragment occurs more than once in the structure then all occurrences of the fragment will be displayed.

9 Output Options

9.1 Saving a Mogul Search

- To save a Mogul search, select **File** from the top-level menu followed by **Save Search...** in the resulting pull-down menu.
- Mogul search files will be given the extension *.mog*.
- Search (*.mog*) files are binary and contain all the information necessary to read the search back into Mogul (select **File** followed by **Open Search...**) and display the search as if had just been run.

Note: All fragment searches cannot be saved in this way. Only the search for the parameter under study will be saved.

9.2 Exporting Entries

Select **File** from the top-level menu followed by **Export Entries...** in the resulting pull-down menu to export the structures currently loaded into the *View structures* pane (you must be in *View structures* to do this).

Available format are:

- Mol2: Tripos Inc. format for 3D molecules and crystal structures.
- PDB: Protein Data Bank format for 3D molecules.
- Refcode: simple ASCII refcode list (default extension *.gcd*) which can be read into ConQuest.

9.3 Exporting Selected Data

- CSD refcodes and geometric parameter values for the currently selected hits (see Viewing a Histogram and Selecting Hits) can be exported.
- Refcodes and values are exported as a *Mogul Hits text file (*.txt)*.
- Select **File** from the top-level menu, then select **Export selected data...** followed by **Refcode and value only...** from the resulting menus**.**
- Alternatively, in the *Results and analysis* pane, click the right-hand mouse button anywhere within the histogram display area, then select **Export selected data** followed by **Refcode and value only...** from the resulting menu.

9.4 Exporting All-Fragments Search Results

- All-Fragments search results are displayed in spreadsheet format. For each fragment type (i.e. bond, angle or torsion) individual fragments are listed in the spreadsheet together with their associated summary statistics.
- To export results for the current fragment type, click on the **Export...** button in the *All-fragments search* window.

- The contents of the *All-fragments search* window are exported as a *Mogul Hits text file (*.txt)*.

9.5 Printing a Histogram

To print the histogram currently displayed within the *Results and analysis* pane, together with its associated summary statistics, either:

- Select **File** from the top-level menu, followed by **Print...** from the resulting drop-down menu.
- Click the right-hand mouse button anywhere within the histogram display area, then select **Print...** from the resulting menu.

To change the display style of a histogram right-click in the histogram display area, and from the resulting menu select the required option (**Font..., Selected Colour..., Deselected Colour..., Background Colour...**).

Mogul User Guide: Instruction File Interface

1 The Mogul Instruction File
2 Format and Example of the Instruction File
3 Mogul Instructions

1 The Mogul Instruction File

1.1 Using an Instruction File with Mogul

A Mogul instruction file allows Mogul to be run automatically in batch mode and provides a way of integrating Mogul with other applications. To start Mogul using an instruction file, issue the following command:

Windows:
```
<INSTALLDIR>\mogul.exe -ins <instruction_file>
```

UNIX:
```
<INSTALLDIR>/bin/mogul -ins <instruction_file>
```

where `<INSTALLDIR>` is the Mogul installation directory and `<instruction_file>` is the name of the Mogul instruction file. For example:

Windows:
```
C:\Program Files\CCDC\Mogul Beta\mogul.exe -ins instructions.txt
```

UNIX:
```
/usr/local/cambridge/mogul/bin/mogul -ins instructions.txt
```

On Windows, it may be possible to identify the location of the most recent version of `mogul.exe` via the Windows registry. The following registry keys will have been written by the Mogul Software installer in either `HKEY_LOCAL_MACHINE` or `HKEY_CURRENT_USER`:

`Software\CCDC\MogulLatestVersion = <ver>` (where `<ver> = 1.2` for example)
`Software\CCDC\Mogul\<ver>\Executable = <INSTALLDIR>\mogul.exe`

1.2 File Locations

Mogul should automatically pick up the locations of data and other files. If it does not then these can be provided via additional command line options as follows:

`-datapath <data_directory>`	Path to Mogul data `<data_directory>` should contain the file `rfactor.idx`

`-csdpath <csd_directory>`	Path to CSD `<csd_directory>` should contain the file `as<version no>be.inf` where `<version no>` depends on year of release
`-sketcher <sketcher_location>`	Path to CCDC Sketcher `<sketcher_location>` should contain the file `lib\sketch.py`

- Mogul requires a valid `<data_directory>` to operate.
- Although Mogul will work without a valid `<csd_directory>`, it will be unable to honour certain search instructions without this; in particular, it will not be able to perform generalised searches or to apply filters.
- Mogul will work without a valid `<sketcher_location>` but some functionality available in the GUI will be missing.

1.3 Default Settings

Default values for some instructions may depend on:

- The Mogul initialisation file
- Settings saved by a user

with settings saved by the user taking precedence over the Mogul initialisation file. Settings provided in a Mogul instruction file will take precedence over both of the above.

When using an instruction file Mogul will by default ***ignore*** settings saved by the user. This is because these may vary from user to user and searches run on different machines by different people may unexpectedly give different results. In some circumstances it may be desirable to have Mogul load user settings on startup when using an instruction file. To make this happen, use the `load_user_settings` command line option. Note that if an instruction file is not used, Mogul will always load a user's saved settings.

The Mogul initialisation file is called `mogul.ini` and is located in:

Windows:
`<INSTALLDIR>`

UNIX:
`<INSTALLDIR>/bin`

Changing settings in this file will impact on the behaviour of Mogul for anyone who runs the installation in `<INSTALLDIR>`.

2 Format and Example of the Instruction File

The instruction file is a plain text file containing Mogul instructions with one instruction per line. Instructions are case insensitive. Typically, the order in which instructions are provided is not important although some commands will supersede any earlier ones relating to the same area of functionality. Any text appearing after a **#** will be treated as a comment and ignored. The following provides an example of a Mogul instruction file that will perform a search for a specified torsion in an input molecule and output results to a file:

```
# Mogul instruction file comment
MOGUL MOLECULE FILE Z:\data\molecules\arachnid.mol2  # comment
MOGUL OUTPUT FILE Z:\data\stats\arachnid.txt
TORSION 2 4 9 3 # search for torsion defined by atom numbers 2, 4, 9 and 3
CONFIG DISTRIBUTION TORSION BIN_WIDTH 15 # write histogram using 15deg bins
CONFIG DISTRIBUTION TORSION BIN_WIDTH 5  # supersedes instruction above
```

3 Mogul Instructions

Mogul instructions fall into the following categories:

- **MOGUL GUI**
  Instructions for controlling the graphical user interface, these include:

  ```
  MOGUL GUI OPEN [ 2D | 3D ]
  ```

- **MOGUL MOLECULE**
  Instructions for specification of input molecule(s), these include:

  ```
  MOGUL MOLECULE FILE <filename>
  MOGUL MOLECULE DIRECTORY <dirname>
  ```

- **MOGUL EDIT**
  Instructions for controlling the changes automatically made to the molecule, these include:

  ```
  MOGUL EDIT BOND_TYPES GUESS <option>
  MOGUL EDIT BOND_TYPES STANDARDISE <option> [ ON | OFF ]
  MOGUL EDIT HYDROGENS GENERATE <option>
  ```

- **BOND, ANGLE, TORSION and RING**
  Instructions for specifying what bonds, angles or torsions to search for, these include:

  ```
  BOND, ANGLE, TORSION or RING atid1 atid2 ... [ <exp_value> ]
  BOND, ANGLE , TORSION or RING ALL
  ```

- **MOGUL OUTPUT**
  Instructions for controlling what statistics are output and to where, these include:

  ```
  MOGUL OUTPUT FILE <filename>
  MOGUL OUTPUT DISTRIBUTION <fragment_type> [ ON | OFF ]
  ```

- **CONFIG SEARCH**
  Instructions relating to searches, these include:

  ```
  CONFIG SEARCH instructions
  CONFIG SEARCH ALL GENERALISATION ON | OFF
  CONFIG SEARCH ALL IMPOSE_UPPER_LEVEL_LIMITS ON | OFF
  ```

```
CONFIG SEARCH <fragment_type> MIN_OBSERVATIONS EXACT <nobs>
CONFIG SEARCH <fragment type> MIN_OBSERVATIONS GENERALISED
<nobs>
CONFIG SEARCH <fragment_type> SELECT <option>
CONFIG SEARCH <fragment_type> SLOW_THRESHOLD <nfrags>
CONFIG SEARCH <fragment_type> MIN_RELEVANCE <r>
CONFIG SEARCH ALL FILTER RFACTOR 0.05 | 0.075 | 0.1 | NONE
CONFIG SEARCH ALL FILTER HEAVIEST_ELEMENT <atomic_number> |
<element_symbol> | NONE
CONFIG SEARCH ALL FILTER EXCLUDE_SOLVENTS | EXCLUDE_NON_SOLVENTS
CONFIG SEARCH ALL FILTER EXCLUDE_ORGANICS |
EXCLUDE_ORGANOMETALLICS
```

- **CONFIG CLASSIFICATION**
  Instructions relating to the classification of fragments as either usual or unusual, these include:

```
CONFIG CLASSIFICATION <fragment_type> UNUSUAL <measure>
<threshold>
CONFIG CLASSIFICATION <fragment_type> FEW_HITS nobs <threshold>
```

- **CONFIG DISTRIBUTION**
  Instructions relating to statistics, these include:

```
CONFIG DISTRIBUTION <fragment_type> BIN_WIDTH <value>
```

- **CONFIG OUTPUT**
  Instructions relating to output, these include:

```
CONFIG OUTPUT FORMAT DEFAULT | TSV | CSV
CONFIG OUTPUT ITEMS item1 [ item2 item3 ... ]
CONFIG OUTPUT INVALID_FRAGMENTS INCLUDE | EXCLUDE
CONFIG OUTPUT MESSAGES <type> ON | OFF
```

3.1 MOGUL GUI OPEN [2D | 3D]

The MOGUL GUI OPEN command controls whether or not the Mogul GUI is opened and in what mode.

MOGUL GUI OPEN 2D
- Opens the GUI in 2D mode (i.e., the molecule is displayed as a two-dimensional object).

MOGUL GUI OPEN 3D
- Opens the GUI in 3D mode (i.e., the molecule is displayed as a three-dimensional object).

MOGUL GUI OPEN
- If no input molecule is specified via MOGUL MOLECULE, this opens the GUI in the default mode (currently 2D).
- If an input molecule is specified, then Mogul will determine the appropriate mode from the input molecule:
  - if any non-zero z-coordinate is detected then the 3D mode will be used; otherwise the 2D mode will be used.

Default Behaviour:

- If no MOGUL GUI OPEN command is specified then the GUI will not be opened except when a molecule is provided via MOGUL MOLECULE but no search instructions are specified (via BOND, ANGLE or TORSION).

Notes:

- If either of the 2D or 3D options are specified and an input molecule is provided via MOGUL MOLECULE then Mogul will use the mode specified to display the input molecule regardless of whether or not this is the most appropriate. For molecules subsequently loaded via the GUI, Mogul will determine the most appropriate mode and use that.
- If the MOGUL GUI OPEN command is specified, along with instructions defining query fragments, then Mogul will search only for the first fragment defined and will automatically display the resulting distribution. If no fragments are specified then Mogul will display the *Build Query* screen, with a molecule loaded, if one is provided via the MOGUL MOLECULE instruction.

3.2 MOGUL MOLECULE FILE <filename>

Except when used in conjunction with MOGUL MOLECULE DIRECTORY, the MOGUL MOLECULE FILE instruction is used to provide Mogul with a query molecule and must be followed by a filename, e.g.

```
MOGUL MOLECULE FILE Z:\data\molecules\arachnid.mol2
```

The file should contain a molecule in any of the formats recognised by Mogul which are:

- cif
- SHELX res
- pdb
- MDL mol
- Tripos mol2
- ConQuest con

Used in conjunction with `MOGUL MOLECULE DIRECTORY`, the `MOGUL MOLECULE FILE` instruction specifies filters that select specific files from a directory, see **`MOGUL MOLECULE DIRECTORY <dirname>`**.

Notes:

- If the file contains concatenated format blocks then Mogul will attempt to process each block unless the GUI is to be opened when it will only process the first block.
- If processing more than one block then:
  - If the default output format is requested then results for the molecule(s) in each block will be preceded by a line of the form

    `MOLECULE <filename> <index> <identifier>`

    where `<index>` is the number of the block in the file (starting at 1) and `<identifier>` is the identifier of the structure it contains if known.
  - For TSV and CSV output, `<filename>`, `<index>` and `<identifier>` will be written to the start of an output line unless the items that are output are otherwise specified by the **`CONFIG OUTPUT ITEMS item1 [ item2 item3 ... ]`** instruction.
  - If Mogul is unable to extract and process a particular block, it will continue to the next block and try to process this. For a block that cannot be processed the word `ERROR` will appear instead of the `<identifier>` on the `MOLECULE` output line.
  - For the default output format, messages about problems encountered when processing a block will be written to the output file after the `MOLECULE` line unless warning messages have been suppressed by the `CONFIG OUTPUT MESSAGES` instruction. For TSV and CSV output, problems will be written only to the console and not to the output file.
- If a format block represents more than one molecule then all molecules in that block will be read.
- If a filename is not provided then Mogul will stop processing the instruction file and exit.

- If `MOGUL MOLECULE FILE` is specified but no search instructions are provided (via `BOND`, `ANGLE` or `TORSION`) then the GUI will automatically open and the molecule(s) in the first format block will be displayed.

3.3 MOGUL MOLECULE DIRECTORY <dirname>

The `MOGUL MOLECULE DIRECTORY` instruction is used to provide Mogul with a directory from which to select input molecule files and must be followed by the name of a directory, e.g.

```
MOGUL MOLECULE DIRECTORY /home/user/mymolecules
```

Mogul will assume that all files in the directory contain molecules and try processing these unless a filter is specified using the `MOGUL MOLECULE FILE` instruction. For example:

```
MOGUL MOLECULE DIRECTORY /home/user/mymolecules
MOGUL MOLECULE FILE *.mol2
```

tells Mogul to only look at files with a .mol2 extension in the directory `/home/user/mymolecules`

More than one filter can be specified using a semi-colon (;) as a separator. For example:

```
MOGUL MOLECULE DIRECTORY /home/user/mymolecules
MOGUL MOLECULE FILE *.mol2;latest_*.sd
```

tells Mogul to look only at files with a .mol2 extension plus files with a .sd extension that have a filename beginning with `"latest_"`.

For details on how Mogul handles individual molecule files, see **`MOGUL MOLECULE FILE <filename>`**.

Notes:

- If Mogul does not recognise a file provided via `MOGUL MOLECULE DIRECTORY` as being in one of the file formats known to it then the file will be ignored.
- If the GUI is to be opened then Mogul will ony process the first molecule in the first file in a recognised format.
- If it is likely that more than one molecule will be processed as a result of the `MOLECULE DIRECTORY` instruction then Mogul will output a `MOLECULE` line to an output file as outlined in `MOGUL MOLECULE FILE`.

3.4 MOGUL EDIT BOND_TYPES GUESS <option>

Mogul is unable to provide results if there are bonds of unknown type in the vicinity of the query fragment. This may well be the case if an input molecule is in a format that does not allow bond types to be specified. Mogul will try and guess the types of bonds in incoming molecules and `MOGUL EDIT BOND_TYPES GUESS` instructions provide control over this.

- `MOGUL EDIT BOND_TYPES GUESS UNKNOWN_3D`
  Requests that Mogul guesses the types of any bonds with unknown type in a 3D molecule.

- `MOGUL EDIT BOND_TYPES GUESS ALL_3D`
  Requests that Mogul guesses the types of all bonds in a 3D molecule even if they are already specified in the input file.

- `MOGUL EDIT BOND_TYPES GUESS NONE`
  Requests that Mogul does not guess any bond types.

Default Behaviour:

- If no `MOGUL EDIT BOND_TYPES GUESS` instruction is provided then Mogul will guess the types of bonds with unknown type.
- This is the same as specifying `MOGUL EDIT BOND_TYPES GUESS UNKNOWN_3D`

Notes:

- Mogul will only guess the types of bonds in a 3D input molecule; a request to guess the types of bonds in a 2D input molecule will be ignored.
- Mogul's bond type guessing is not perfect. If a query fragment unexpectedly gives no results it might be worth checking any bond types that Mogul has automatically assigned.

3.5 MOGUL EDIT BOND_TYPES STANDARDISE <option> [ON | OFF]

If the bond types of an input molecule do not conform to Cambridge Structural Database (CSD) conventions then Mogul will give incorrect results. For example, a CSD convention is to code a benzene ring with 6 C-C bonds of aromatic type and not with alternating single and double bonds. Mogul will attempt to standardise the bond types of incoming molecules and the `MOGUL EDIT BOND_TYPES STANDARDISE` instruction provides control over this.

- `MOGUL EDIT BOND_TYPES STANDARDISE AROMATIC [ ON | OFF ]`
  - If `ON` is specified then Mogul will identify and set aromatic bonds according to CSD conventions.

- This option is `ON` by default unless suppressed by specifying `OFF`

- `MOGUL EDIT BOND_TYPES STANDARDISE DELOCALISED [ ON | OFF ]`
  - If `ON` is specified then Mogul will identify and set delocalised bonds according to CSD conventions.
  - This option is `ON` by default unless suppressed by specifying `OFF`

- `MOGUL EDIT BOND_TYPES STANDARDISE ALL [ ON | OFF ]`
  - This instruction allows all of the above options to be turned `ON` or `OFF`

Default Behaviour:

- If no `MOGUL EDIT BOND_TYPES STANDARDISE` instruction is provided then Mogul will standardise aromatic and delocalised bonds.
- This is the same as specifying `MOGUL EDIT BOND_TYPES STANDARDISE ALL ON`

Notes:

- If neither `ON` or `OFF` are specified then this is the same as specifying `ON`.
- Standardisation can be applied to either 2D or 3D molecules.
- Mogul cannot standardise bond types if the input molecule contains bonds of unknown type or atoms of unknown element type. If these are detected then standardisation instructions will be ignored.
- However, if the input molecule contains unknown bonds, but is 3D and a `MOGUL EDIT BOND_TYPES GUESS UNKNOWN_3D` instruction is specified, then standardisation will be performed provided an appropriate standardisation instruction is included.
- If the `MOGUL EDIT BOND_TYPES GUESS ALL_3D` instruction is specified then standardisation is carried out automatically.
- If no option follows `STANDARDISE` then Mogul will stop processing the instruction file and exit.
- Mogul's bond type standardisation is not perfect. If a query fragment unexpectedly gives no results it might be worth checking any bond type standardisation Mogul has applied.

3.6 `MOGUL EDIT HYDROGENS GENERATE <option>`

If hydrogens are missing from an input molecule then Mogul may give incorrect results. Mogul will thus attempt to add missing hydrogens. The `MOGUL EDIT HYDROGENS GENERATE` instruction provides control over this.

- `MOGUL EDIT HYDROGENS GENERATE MISSING`

- This will add missing hydrogens but preserve those already present.

- `MOGUL EDIT HYDROGENS GENERATE ALL`
  - This will remove any hydrogens that are present and add those Mogul believes should be present.
- `MOGUL EDIT HYDROGENS GENERATE NONE`
  - This will prevent Mogul from adding any hydrogens.

Default Behaviour:

- If no `MOGUL EDIT HYDROGENS GENERATE` instruction is provided then:
  - if the input molecule contains *no* hydrogens, `MOGUL EDIT HYDROGENS GENERATE MISSING` is assumed and Mogul will add all hydrogens that it determines should be present;
  - if the input molecule contains *any* hydrogens, `MOGUL EDIT HYDROGENS GENERATE NONE` is assumed and no hydrogens will be added.

Notes:

- The number of hydrogens added to an atom may be affected by the atom's formal charge.
- Some file formats do not store atomic charge, in which case all charges will be assumed to be zero. Mogul may therefore add an incorrect number of hydrogens to any atoms that should, in fact, be charged.
- Mogul cannot determine the number of hydrogens to add if a molecule contains bonds of unknown type or atoms of unknown element type. If these are detected then a request to add hydrogens will be ignored. However, hydrogens can be added to a 3D molecule with unknown bonds if one of the instructions `MOGUL EDIT BOND_TYPES GUESS UNKNOWN_3D` or `MOGUL EDIT BOND_TYPES GUESS ALL_3D` is specified.
- If no option follows `GENERATE` then Mogul will stop processing the instruction file and exit.
- Addition of hydrogens by Mogul will not always be completely correct. If a query fragment unexpectedly gives no results it might be worth checking that any hydrogens Mogul has added are correct.

3.7 BOND, ANGLE, TORSION or RING atid1 atid2 ... [<exp_value>]

```
BOND atid1 atid2 [ <exp_value> ]
ANGLE atid1 atid2 atid3 [ <exp_value> ]
TORSION atid1 atid2 atid3 atid4 [ <exp_value> ]
RING atid1 atid2 atid3 atid4 atid 5 [atid6 ....] [ <exp_value> ]
```

The `BOND`, `ANGLE`, `TORSION` and `RING` instructions can be used to request a search for a specific fragment.

Mogul will preserve the order of atoms in an input molecule file and so fragments should be defined by specifying the integer index of the atom in the file. Thus, for a molecule with 11 atoms:

- the first atom listed in the file should be identified as 1;
- the last atom listed in the file should be identified as 11.

Given a CIF containing the following atoms:

```
S1 S 0.72975(3) 0.47834(8) 0.34345(3)
S2 S 0.92926(3) 0.24748(7) 0.38691(3)
N1 N 0.85667(10) 0.6395(2) 0.47564(10)
H1 H 0.9123(16) 0.654(4) 0.5156(16)
C1 C 0.84573(11) 0.4634(3) 0.40933(11)
C2 C 0.76952(11) 0.7892(3) 0.49291(12)
H2 H 0.78910 0.96260 0.50410
H3 H 0.73790 0.72900 0.55130
C3 C 0.69813(12) 0.7663(3) 0.40126(13)
H4 H 0.70640 0.90590 0.35630
H5 H 0.62840 0.76320 0.41890
```

the fragment `C1-S1-C3` would therefore be defined as:

```
ANGLE 5 1 9
```

Atom indices may be optionally followed by the actual value of the bond length, valence angle or torsion angle in the input molecule, for example:

```
ANGLE 5 1 9  92.81
```

Mogul does not currently use this value but it will be included in the instruction line written back to the output file.

3.8 BOND, ANGLE , TORSION or RING ALL

```
BOND ALL
ANGLE ALL
TORSION ALL
RING ALL
```

Instead of specifying specific fragments via atom indices, Mogul can be asked to perform searches for all fragments of a particular type by following the `BOND`, `ANGLE` or `TORSION` instructions with `ALL`.

Notes:

- Currently, an `ALL` instruction will be ignored if the `MOGUL GUI OPEN` instruction is present.
- Mogul will not perform duplicate searches for any fragments defined via atom indices if they will be searched for as the result of an `ALL` instruction.
- By default, Mogul will not generate output for any fragment that is found to be invalid when performing `ALL` searches. This behaviour can be overridden via the `CONFIG SEARCH` instruction.

3.9 `MOGUL OUTPUT FILE <filename>`

There are a number character sequences that Mogul will replace with information derived from an input molecule file. These are:

%d : directory containing the molecule file
%f : file root (or base) without extension
%e : extension of molecule filename
%n : name of molecule in file
%i : index of molecule in file
%c : count of molecules processed maintained by Mogul during a run

By using these it is possible to direct output to separate files when more than one input molecule or input file is processed during a run.

For example, if Mogul is to process 2 files first.sd and second.sd then:

```
MOGUL OUTPUT FILE %d/output/%f.txt
```

will create output files first.txt and second.txt in a subdirectory of the directory containing the molecule files called output. Note that Mogul will not create any missing subdirectories and will thus not generate any output files if sub-directories are not present.

If each file contained 2 molecules then:

```
MOGUL OUTPUT FILE %d/rank_%i.txt
```

will create files rank_1.txt and rank_2.txt in the directory containing the input files. rank_1.txt will contain output for the two molecules that are found first in each of the two input files and rank_2.txt the second.

If the 4 molecules in the 2 files each have unique names MOL1, MOL2, MOL3 and MOL4 then:

```
MOGUL OUTPUT FILE %n.txt
```

will create file MOL1.txt, MOL2.txt, MOL3.txt and MOL4.txt with each containing just the results for one molecule. Note that if Mogul cannot determine a molecule name it will make one up and if searches involve more than one file, more than one unnamed molecule may end up with the same name.

To be sure of getting a unique filename for each molecule then either use sufficient identifiers to distinguish each molecule or use %c which will result in insertion of an arbitrary but unique integer for each molecule processed.

3.10 MOGUL OUTPUT DISTRIBUTION <fragment_type> [ON | OFF]

The `MOGUL OUTPUT DISTRIBUTION` instruction will result in a representation of the Mogul distribution being written to the Mogul output file.

- `<fragment_type>` must be one of `BOND`, `ANGLE`, `TORSION` or `ALL` and may be followed by either `ON` or `OFF`. For example:

  ```
  MOGUL OUTPUT DISTRIBUTION BOND ON # output distributions for bonds
  MOGUL OUTPUT DISTRIBUTION TORSION OFF # suppress distributions for torsions
  ```

- If neither `ON` or `OFF` are specified then this is the equivalent of specifying `ON`.

- The distribution is represented in the output file by a line such as:

  ```
  DISTRIBUTION 0 180 10 18 : 279 78 20 0 1 0 0 0 0 0 0 0 0 1 0 19 80 278
  ```

- The four numbers preceding the colon (`:`) indicate:
  - lower bound of first bin;
  - upper bound of last bin;
  - bin size;
  - number of bins.

- The numbers following the colon (`:`) indicate how many observations there are in each bin.

- With the exception of torsions, Mogul will only output as many bins as are necessary to represent all observations.

    - For torsions, bins always run from 0 to 180 degrees (the sign of a torsion angle is ignored in Mogul).

- Default bin sizes are:
    - bonds 0.01Å
    - angles 0.25^{o}
    - torsion 10^{o}

The bin size can be configured via the `CONFIG DISTRIBUTION` instruction.

A bin *lower* to *upper* will include values equal to *lower*. It will not include values equal to *upper*: these will be included in the next bin except for the last bin where values equal to *upper* will be included as well as values equal to *lower*.

Default Behaviour:

- If no `MOGUL OUTPUT DISTRIBUTION` instructions are provided then, by default, a distribution will be output for torsions but not for angles or bonds.

3.11 `CONFIG SEARCH` instructions

There are a number of instructions for specifying program settings that control the speed and quality of Mogul searches. In order to explain these, it is necessary to define some Mogul concepts:

- A Mogul *fragment* includes the atoms that define a bond, angle, torsion or ring, and their bonded neighbours.
- The *query fragment* is the fragment that is being searched for.
- A *hit fragment* is a CSD fragment found by a Mogul search on a query fragment.
- An *exact distribution* is one where all the hit fragments are identical to the query fragment.
- A *generalised distribution* is one where at least some hit fragments are similar but not identical to the query fragment.
- An *exact search* is one that returns an exact distribution.
- A *generalised search* is one that returns a generalised distribution.
- The *relevance* of a hit fragment is a measure of how closely related it is to the query fragment. Relevance values range from 0.0 to 1.0. A hit fragment will have a relevance of 1.0 if and only if it is identical to the query fragment.

Configurable search settings can be used to control:

- The minimum number of observations that a search should try to find.
- Whether a generalised distribution is to be returned if there are insufficient observations in the exact distribution.
- The minimum acceptable relevance for a hit fragment in a generalised distribution.
- How a generalised search is to be conducted: in particular, the balance between the speed of the search and the relevance of the results.
- Whether hit fragments should be rejected if they come from structures with high R-factors.

General Notes:

- In the following instructions, `<fragment_type>` must be one of `BOND, ANGLE, TORSION` or `RING`.
- Settings provided via the following instructions will be reflected in the GUI if opening this via an instruction file (with the exception of `CONFIG SEARCH ALL_FRAGS INCLUDE_INVALID`).
- Default values for most settings are provided in the Mogul initialisation file. This is an ASCII file called `mogul.ini` located in the same directory as the Mogul executable (see Default Settings).

3.12 CONFIG SEARCH ALL GENERALISATION ON | OFF

This instruction tells Mogul whether or not to automatically perform generalised searches.

If `OFF` is specified then Mogul will perform an exact search and only return exact distributions even if other `CONFIG SEARCH` criteria are not satisfied.

If `ON` is specified then Mogul will perform generalised searches according to the criteria set by the following instructions:

```
CONFIG SEARCH <fragment_type> MIN_OBSERVATIONS EXACT
CONFIG SEARCH <fragment_type> MIN_OBSERVATIONS GENERALISED
CONFIG SEARCH <fragment_type> SELECT
CONFIG SEARCH <fragment_type> MIN_RELEVANCE
```

Default Behaviour:

- If no `GENERALISATION` instruction is provided then Mogul will perform generalised searches (the equivalent of `CONFIG SEARCH ALL GENERALISATION ON`) unless this is overriden in the Mogul Initialisation file or by a users saved settings.

3.13 CONFIG SEARCH ALL IMPOSE_UPPER_LEVEL_LIMITS ON | OFF

This instruction tells Mogul whether or not to limit the number of levels traversed for generlaised searches. Occasionally Mogul can take a very long time to identify similar fragments when performing a generalised search. Limiting the number of levels traversed will reduce the chances of this happening but may also result in fewer hits being found.

If `OFF` is specified then Mogul will, for generalised searches, proceed up the tree until the required number of hits are found that satisfy the relevance criteria.

If `ON` is specified then Mogul will only traverse a set number of levels.

Default Behaviour:

- If no `IMPOSE_UPPER_LEVEL_LIMITS` instruction is provided then Mogul will perform generalised searches without imposing an upper level limits (the equivalent of `CONFIG SEARCH ALL IMPOSE_UPPER_LEVEL_LIMITS ON`) unless this is overriden in the Mogul Initialisation file or by a users saved settings.

3.14 CONFIG SEARCH <fragment_type> MIN_OBSERVATIONS EXACT <nobs>

`MIN_OBSERVATIONS EXACT` specifies the minimum acceptable size of an exact distribution. If there is not a distribution containing at least `<nobs>` fragments identical to the query fragment then Mogul will perform a generalised search according to the criteria specified by other `CONFIG SEARCH` instructions.

Default Behaviour:

- If no `MIN_OBSERVATIONS EXACT` instruction is specified for a particular fragment type then the default behaviour will be dictated by settings in the Mogul Initialisation File or saved user settings (see Default Settings). CCDC default values for this instruction are:
  - bonds 15
  - angles 15
  - torsion 40
  - rings 15

Notes:

- `<nobs>` must be greater than or equal to zero.
- If `<nobs>` is zero, Mogul will not perform a generalised search even if the exact distribution contains no observations.

3.15 CONFIG SEARCH <fragment type> MIN_OBSERVATIONS GENERALISED

<nobs>

MIN_OBSERVATIONS GENERALISED specifies the minimum number of observations that Mogul should try to find when performing a generalised search.

Default Behaviour:

- If no MIN_OBSERVATIONS GENERALISED instruction is specified for a particular fragment type then the default behaviour will be dictated by settings in the Mogul Initialisation File or saved user settings (see Default Settings). CCDC default values for this instruction are:
  - bonds 15
  - angles 15
  - torsion 40
  - rings 15

Notes:

- <nobs> must be greater than or equal to zero.
- If <nobs> is zero, Mogul will not perform a generalised search.
- The number of observations in the resulting distribution may sometimes be less than the number specified by <nobs>. This will happen, e.g., if there are not enough fragments that satisfy other criteria used to control the search.

3.16 CONFIG SEARCH <fragment_type> SELECT <option>

As a generalised search progresses, Mogul identifies fragments that may satisfy the various search criteria. The SELECT instruction dictates which fragments are included in the final generalised distribution. It provides control over the following factors:

- the size of the distribution
- the relevance of the hit fragments in the distribution
- the speed of the search

The possible values for <option> are:

- BEST: This will try to find the most relevant fragments possible but may be slow. The size of the resulting distribution will usually be close to the value specified by MIN_OBSERVATIONS GENERALISED, though this is not guaranteed.
- ANY: This gives the fastest search speeds but will probably not find the most relevant fragments possible. However, all hit fragments will have a relevance of at least MIN_RELEVANCE and if there are any fragments in the Mogul library that are identical to the query (relevance = 1.0),

they are guaranteed to be included in the distribution, provided they satisfy other search criteria such as FILTER RFACTOR. The size of the distribution will usually be close to the value specified by MIN_OBSERVATIONS GENERALISED, though this is not guaranteed.

- BEST_UNLESS_SLOW: This is a compromise between SELECT BEST and SELECT ANY. The distribution will be identical to that which would have been produced by SELECT BEST unless this would necessitate Mogul looking at a large number of fragments, which would cause the search to be slow. In this case, the selection mode will switch from SELECT BEST to SELECT ANY. See **CONFIG SEARCH <fragment_type> SLOW_THRESHOLD <nfrags>** for details on how to customise this option.
- ALL: This will include in the distribution all fragments whose relevance is at least MIN_RELEVANCE, provided they satisfy other search criteria such as FILTER RFACTOR. This may result in a large distribution containing a lot of fragments. The search may also take a long time.

Default Behaviour:

- If no SELECT instruction is specified for a particular fragment type then the default behaviour will be dictated by settings in the Mogul initialisation file or saved user settings (see Default Settings). CCDC default options for this instruction are:
  - bonds BEST
  - angles BEST
  - torsion BEST
  - rings BEST

3.17 CONFIG SEARCH <fragment_type> SLOW_THRESHOLD <nfrags>

There are several stages to a generalised search and at each stage there are a number of distinct fragments for Mogul to examine. If the final distribution is to include the most relevant fragments (i.e SELECT BEST) then Mogul must examine all distinct fragments identified at each stage of the search. The greater the number of these, the longer the search will take.

The SLOW_THRESHOLD instruction controls the behaviour of the SELECT BEST_UNLESS_SLOW option. If using SELECT BEST_UNLESS_SLOW then Mogul will set out to identify the most relevant fragments, i.e. as if SELECT BEST had been specified. If, at any point, Mogul is presented with a number of distinct fragments to examine that exceeds the value <nfrags> (as specified by the SLOW_THRESHOLD instruction), it will switch to a SELECT ANY mode of operation, i.e. from this point, it will take any fragment whose relevance is at least MIN_RELEVANCE.

Default Behaviour:

- If no SLOW_THRESHOLD instruction is specified for a particular fragment type then the default

behaviour will be dictated by settings in the Mogul Initialisation File or saved user settings (see Default Settings). CCDC default options for this instruction are:

- bonds 500
- angles 500
- torsion 500
- rings 500

Notes:

- The `SLOW_THRESHOLD` instruction affects only `SELECT BEST_UNLESS_SLOW`. It has no effect on other `SELECT` options.
- The value of `<nfrags>` must be greater than or equal to zero. If it is zero and `SELECT BEST_UNLESS_SLOW` is specified then this is equivalent to specifying `SELECT ANY`.

3.18 CONFIG SEARCH <fragment_type> MIN_RELEVANCE <r>

Mogul determines how similar a fragment is to the query by calculating a relevance value. The `MIN_RELEVANCE` instruction tells Mogul to accept in a generalised search only fragments whose relevance is at least `<r>`.

Default Behaviour:

- If no `MIN_RELEVANCE` instruction is specified for a particular fragment type then the default behaviour will be dictated by settings in the Mogul initialisation file or saved user settings (see Default Settings). CCDC default options for this instruction are:
  - bonds 0.75
  - angles 0.75
  - torsion 0.75
  - rings 0.75

Notes:

- The value of `<r>` must be in the range 0.0 to 1.0.
- If a fragment is identical to the query fragment it will have a relevance of 1.0; otherwise it will have a relevance of less than 1.0.
- Although Mogul allows `<r>` to be any value in the range 0.0 - 1.0, it is recommended that a value of 0.75 or greater is used.
- If Mogul cannot find enough fragments with a relevance of at least `<r>` then the size of a generalised distribution may be less than that specified by `MIN_OBSERVATIONS GENERALISED`.

3.19 CONFIG SEARCH ALL FILTER RFACTOR 0.05 | 0.075 | 0.1 | NONE

The FILTER RFACTOR instruction tells Mogul to accept only fragments from CSD structures with R-factor below a specified value.

The required FILTER RFACTOR value is expressed as a decimal number (not as a percentage) and represents the maximum allowed R-factor. It can be only one of three numerical values (0.05, 0.075 or 0.1) or NONE (which means fragments will be included regardless of R-factor). Specifying 0.075 (for example) will require that all fragments are from structures with R-factor no higher than 0.075 (or 7.5%).

Default Behaviour:

- If no FILTER RFACTOR instruction is specified then this is the equivalent of specifying FILTER RFACTOR NONE.

Notes:

- If excluding fragments from an exact distribution on the basis of R-factor brings the number of observations below the value specified by MIN_OBSERVATIONS EXACT then Mogul will attempt to find a suitable generalised distribution.
- If Mogul cannot find enough fragments from structures satisfying the specified R-factor limit, then the size of a generalised distribution may be less than that specified by MIN_OBSERVATIONS GENERALISED.
- FILTER RFACTOR applies to all fragments regardless of type; it is not possible to specify different R-factor requirements for different fragment types.

3.20CONFIG SEARCH ALL FILTER HEAVIEST_ELEMENT <atomic_number> | <element_symbol> | NONE

The FILTER HEAVIEST_ELEMENT instruction tells Mogul to ignore fragments from CSD structures that have elements heavier than a specified atomic number or element.

FILTER HEAVIEST_ELEMENT must be followed by either an integer or element symbol representing the heaviest element that is permitted in a structure or NONE to indicate that filtering on heaviest element is not required.

Default Behaviour:

- If no FILTER HEAVIEST_ELEMENT instruction is specified then this is the equivalent of specifying FILTER FILTER HEAVIEST_ELEMENT NONE.

Notes:

- If excluding fragments from an exact distribution on the basis of heaviest element brings the number of observations below the value specified by `MIN_OBSERVATIONS EXACT` then Mogul will attempt to find a suitable generalised distribution.
- If Mogul cannot find enough fragments from structures satisfying the specified heaviest element limit, then the size of a generalised distribution may be less than that specified by `MIN_OBSERVATIONS GENERALISED`.
- `FILTER HEAVIEST_ELEMENT` applies to all fragments regardless of type; it is not possible to specify different heaviest element requirements for different fragment types.

3.21CONFIG SEARCH ALL FILTER EXCLUDE_SOLVENTS | EXCLUDE_NON_SOLVENTS

`FILTER EXCLUDE_SOLVENTS` and `FILTER EXCLUDE_NON_SOLVENTS` instructs Mogul to ignore fragments depending on whether they are from solvent or non- solvent molecules.

If the `FILTER` instruction is followed by:

- `EXCLUDE_SOLVENTS` then Mogul will filter out fragments from solvent molecules
- `EXCLUDE_NON_SOLVENTS` then Mogul will filter out fragments from non-solvent molecules, i.e., only fragments from solvent molecules will be included in the resulting distribution

Default Behaviour:

- If neither `FILTER EXCLUDE_SOLVENTS` nor `FILTER EXCLUDE_NON_SOLVENTS` are specified, Mogul will include fragments from both solvent and non-solvent molecules.

Notes:

- A molecule is considered to be a solvent if it is found in a CCDC catalog of known solvent molecules. However, if a crystal structure contains only solvent molecules then the largest molecule will be assumed to be a non-solvent.
- If both `FILTER EXCLUDE_SOLVENTS` and `FILTER EXCLUDE_NON_SOLVENTS` are included in an instruction file then the instruction that appears later in the file will be used.
- If excluding fragments from an exact distribution on the basis of `EXCLUDE_SOLVENTS` or `EXCLUDE_NON_SOLVENTS` brings the number of observations below the value specified by `MIN_OBSERVATIONS EXACT` then Mogul will attempt to find a suitable generalised distribution.
- If Mogul cannot find enough fragments from structures satisfying the specified `EXCLUDE_SOLVENTS` or `EXCLUDE_NON_SOLVENTS` restriction, then the size of a generalised distribution may be less than that specified by `MIN_OBSERVATIONS GENERALISED`.
- `EXCLUDE_SOLVENTS` and `EXCLUDE_NON_SOLVENTS` applies to all fragments regardless

of type; it is not possible to specify different heaviest element requirements for different fragment types.

3.22CONFIG SEARCH ALL FILTER EXCLUDE_ORGANICS | EXCLUDE_ORGANOMETALLICS

`FILTER EXCLUDE_ORGANICS` and `FILTER EXCLUDE_ORGANOMETALLICS` can be used to instruct Mogul to ignore fragments depending on whether they are from organic or organometallic structures.

If the `FILTER` instruction is followed by:

- `EXCLUDE_ORGANICS` then Mogul will filter out fragments from organic molecules, i.e., only fragments from organometallic structures will be included in the resulting distribution
- `EXCLUDE_ORGANOMETALLICS` then Mogul will filter out fragments from organometallic molecules, i.e., only fragments from organic structures will be included in the resulting distribution

***Default Behaviour*:**

- If neither `FILTER EXCLUDE_ORGANICS` nor `FILTER EXCLUDE_ORGANOMETALLICS` are specified, Mogul will include fragments from both organic and organometallic molecules

Notes:

- An organometallic structure is considered to be one containing at least one transition metal, lanthanide, actinide, or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po.
- If both `FILTER EXCLUDE_ORGANICS` and `FILTER EXCLUDE_ORGANOMETALLICS` are included in an instruction file then the instruction that appears later in the file will be used.
- If excluding fragments from an exact distribution on the basis of `EXCLUDE_ORGANICS` or `EXCLUDE_ORGANOMETALLICS` brings the number of observations below the value specified by `MIN_OBSERVATIONS EXACT` then Mogul will attempt to find a suitable generalised distribution.
- If Mogul cannot find enough fragments from structures satisfying the specified `EXCLUDE_ORGANICS` or `EXCLUDE_ORGANOMETALLICS` restriction, then the size of a generalised distribution may be less than that specified by `MIN_OBSERVATIONS GENERALISED`.
- `EXCLUDE_ORGANICS` and `EXCLUDE_ORGANOMETALLICS` applies to all fragments regardless of type; it is not possible to specify different heaviest element requirements for different fragment types.

3.23 CONFIG CLASSIFICATION <fragment_type> UNUSUAL <measure> <thresh-

old>

- <fragment_type> : one of `BOND, ANGLE, TORSION or RING`
- <measure> : one of:
  - dmin : (see Summary Statistics)
  - z-score : (see Summary Statistics) (irrelevant for torsions and rings)
  - mean-x : absolute difference between mean and query value (irrelevant for torsions)
  - local_density (irrelevant for bonds and angles) : (see below)

local_density

- percentage of observed values within x units of query value
- x is defined by `CONFIG DISTRIBUTION <fragtype> LOCAL_DENSITY_TOLERANCE <value>`
- default values of x if `LOCAL_DENSITY_TOLERANCE` is not specified are:
  - torsions: 10 degrees
  - angles: 1 degree
  - bonds: 0.01 Angstrom
  - rings 10 degrees
- It is possible to use optional WITHIN <Interval> in the context of local density if a default value of x is not required e.g. as in `CONFIG CLASSIFICATION RING UNUSUAL local_density 5 WITHIN 20`. This will class as unusual any distributions where less than 5% of the distribution is within 20 degrees of the query ring RMSD (i.e. RMSD <20 degrees)

Notes:

- If this instruction is not provided then fragments will be classified as unusual using:
  - torsions: local_density with threshold 5, i.e., less than 5% of observations are within 10 degrees of query value
  - bonds and angles: z-score with threshold 2.0, i.e., z-score is greater than 2.0

3.24 CONFIG CLASSIFICATION <fragment_type> FEW_HITS nobs <threshold>

- <fragtype> : one of `BOND, ANGLE, TORSION or RING`
- <threshold> : number of hits below which a distribution is considered to have too few hits

Notes:

- Default thresholds if this instruction is not specified are:
  - torsions: 15
  - bonds and angles: 5

3.25 CONFIG DISTRIBUTION <fragment_type> BIN_WIDTH <value>

The `CONFIG DISTRIBUTION` instruction allows control over the width of bins used to represent a distribution output by Mogul.

Default bin widths for distributions are:

- bonds 0.01Å
- angles 0.25^{o}
- torsion 10^{o}

The following instruction will result in a bin width of 5^{o} for torsions, overriding the default:

```
CONFIG DISTRIBUTION TORSION BIN_SIZE 5
```

Notes:

- This instruction will not currently affect the bin width used when displaying distributions as histograms in the Mogul GUI. It will only affect output to a file specified via `MOGUL OUTPUT FILE`.

3.26CONFIG OUTPUT FORMAT DEFAULT | TSV | CSV

The `CONFIG OUTPUT FORMAT` instruction allows control over the format in which statistics and other information is written to the file(s) specified by **`MOGUL OUTPUT FILE <filename>`**.

If `DEFAULT` is specified then this will output data in the following format:

- The output file content may begin with:
  - Errors, preceded by the word `ERROR`
  - Warnings, preceded by the word `WARN`
  - Information (usually relating to changed bond types and addition of hydrogens) preceded by the word `INFO`

- *Note:* Output of WARN and INFO messages can be controlled by the **CONFIG OUTPUT MESSAGES <type> ON | OFF** instruction.

- The output file will include statistics relating to any fragments that are specified in the instruction file, as follows:
  - Each set of results is preceded by the instruction line defining the query fragment.
  - This will be followed by either:
    a line beginning with ERROR if the fragment was found to be invalid;
    a line beginning with NOHITS if there were no results for the fragment;
    a line beginning with STATS followed by basic statistics if the fragment generated results.

- Basic statistics (STATS) for bonds and angles include the following items in the order given:
  number of observations;
  mean;
  minimum;
  maximum;
  median;
  sample deviation;
  upper quartile;
  lower quartile.
- For torsions, basic statistics (STATS) include only the number of observations.
- The STATS line may optionally be followed by a representation of the Mogul distribution beginning with the word DISTRIBUTION. See **MOGUL OUTPUT DISTRIBUTION <fragment_type> [ON | OFF]** for further details. By default, a DISTRIBUTION line will be generated for all torsion fragments that give results.

- An example output file is given below:

```
INFO BOND_TYPES GUESS: Unknown: 0/21; Changed: 0
INFO STANDARDISE AROMATIC: Before: 0; After: 6
INFO HYDROGENS ADD: H already present; assuming these are correct
INFO HYDROGENS ADD: Before: 8; After: 8
BOND 12 24 # invalid fragment
ERROR Invalid Fragment - Error, cannot make fragment. Some atoms
are not bonded
BOND 56 28 1.34239 # fragment giving no hits
NOHITS No hits were found for this query fragment.
ANGLE 8 11 14 119.24570 # valence angle - with stats
STATS 749 120.764 103.329 133.089 120.543 2.11255 121.541
```

```
119.862
TORSION 3 2 10 11 164.923 # torsion angle - stats and
distribution
STATS 305
DISTRIBUTION 0 180 10 18 : 0 0 1 0 0 2 14 37 59 56 34 43 38 18 1
0 1 1
```

If TSV is specified then data is written on a single line with each item separated by a tab.

If CSV is specified then data is written on a single line with each item separated by a comma.

The actual items written and the order in which they are written can be controlled by `CONFIG OUTPUT ITEMS` unless `DEFAULT` format is requested. See **`CONFIG OUTPUT ITEMS item1 [ item2 item3 ... ]`** for further information about the content of a TSV or CSV file.

Notes:

- If no `CONFIG OUTPUT FORMAT` instruction is supplied then this is the same as specifying `CONFIG OUTPUT DEFAULT`.
- For TSV and CSV output, if a value is not available (e.g. some statistics when there are no hits, mean for torsions) then an empty field will be output.
- If `MOGUL OUTPUT DISTRIBUTION` is used in conjunction with TSV and CSV then distribution data will be written at the end of the line using the relevant separator to separate values. Note that this can result in a variable number of columns, particularly when outputting distributions for bonds and angles. By default, a distribution will be output for torsions but not angles and bonds.
- `INFO` and `WARNING` messages are not written to a TSV or CSV files but are dislayed in the console. All console output can be captured in a log file using the "`-logfile <filename>`" command line option when starting Mogul.
- If Mogul encounters a molecule or fragment that cannot be processed then no output will be written to a TSV or CSV file for this molecule or fragment.
- By default, Mogul will output a header line for CSV and TSV files. This can be suppressed using the command `CONFIG OUTPUT HEADER OFF`.

3.27CONFIG OUTPUT ITEMS item1 [item2 item3 ...]

The `CONFIG OUTPUT ITEMS` instruction can be used to control which items are included in TSV and CSV files and in what order. It has no impact if the default output format is requested.

Valid items are:

- Fragment information:
  - molecule_file: name of input molecule file
  - molecule_index: index of molecule in file
  - molecule_name: name of molecule in file
  - fragment_id: output unique fragment identifier
  - fragment_type: type of fragment (BOND, ANGLE or TORSION)
  - atom_indices: index of atoms defining query fragment (relates to order of atoms in file)
  - atom_labels: labels of atoms defining query fragment
  - query_value: bond length, angle or torsion of feature in query molecule (if available)

- Statistics from Mogul:
  - nhits: number of hits
  - mean: mean
  - sd: standard deviation
  - min: minimum value
  - lq: lower quartile
  - median: median
  - uq: upper quertile
  - max: maximum
  - classification: output classification (see below)
  - output of figures of merit:
    local_density
    dmin
    z-score
    mean-x

Default behaviour:

- If no `CONFIG OUTPUT ITEMS` instruction is given then all of the items listed above will be output to a TSV or CSV file in the order listed.

Notes:

- If items are requested but are not available or valid for a particular fragment then an empty field will be output.
- If `CONFIG OUTPUT ITEMS` is specified with no `MOGUL OUTPUT DISTRIBUTION` instructions then the torsion distribution will be appended to the end of each line for torsion fragments. To suppress torsion distributions or also include bond and angle distributions, see the

MOGUL OUTPUT DISTRIBUTION <fragment_type> [ON | OFF] instruction.

- classification:
  If there are hits, will be of the form `<class> (<qualification>)` where:
  - `<class>` is one of Unusual or Not unusual
  - `<qualification>` is one of Enough hits or Few hits
  - Examples: Unusual (Enough hits), Not unusual (Few hits)
    If there are no hits will be just No hits

3.28 CONFIG OUTPUT INVALID_FRAGMENTS INCLUDE | EXCLUDE

This instruction allows control over output generated by Mogul for invalid fragments. Invalid fragments might include:

- Fragments involving metals.
- Cyclic torsions.
- Fragments involving bonds of unknown type or atoms with unknown element type.

If `CONFIG OUTPUT INVALID_FRAGMENTS INCLUDE` is specified then Mogul will generate output for invalid fragments and this will be written to the output file as (e.g.):

```
ANGLE 2 1 3 98.410 # P1 Cd1 S1
ERROR Invalid Fragment - Not in the library.
ANGLE 25 26 27 123.133 # C19 C20 C21
ERROR Invalid Fragment - Unknown atom or bondtype found
```

If `CONFIG OUTPUT INVALID_FRAGMENTS EXCLUDE` is specified then Mogul will suppress output for invalid fragments. Neither the ANGLE lines nor the ERROR lines in the above example will appear in the output file.

Default Behaviour:

- If no `CONFIG OUTPUT INVALID_FRAGMENTS` instruction is provided then Mogul will by default generate output for invalid fragments as well as valid ones.
- This is the same as specifying `CONFIG OUTPUT INVALID_FRAGMENTS INCLUDE`

3.29 CONFIG OUTPUT MESSAGES <type> ON | OFF

The following messages may be written to the Mogul output file:

- Warnings, preceded by the word `WARN`
- Information (usually relating to changed bond types and addition of hydrogens) preceded by the word `INFO`

The `CONFIG OUTPUT MESSAGES` instruction controls output of these:

- `CONFIG OUTPUT MESSAGES INFO ON | OFF`
  - If `ON` is specified then `INFO` messages will be included in the Mogul output file
  - This option is `ON` by default unless suppressed by specifying `OFF`

- `CONFIG OUTPUT MESSAGES WARN ON | OFF`
  - If `ON` is specified then `WARN` messages will be included in the Mogul output file
  - This option is `ON` by default unless suppressed by specifying `OFF`

- `CONFIG OUTPUT MESSAGES ALL ON | OFF`
  - This instruction allows all of the above options to be turned `ON` or `OFF`

Default Behaviour:

- If no `CONFIG OUTPUT MESSAGES` instruction is provided then Mogul will output all messages.
- This is the same as specifying `CONFIG OUTPUT MESSAGES ALL ON`

Notes:

- These instructions only control output to the Mogul output file and will not change what is written to (e.g.) a UNIX console window.
- Output to the UNIX console can be suppressed by issuing the following instruction:

```
CONFIG OUTPUT CONSOLE OFF
```

- This will suppress `INFO` and `WARN` output sent to the console when Mogul is executing an instruction file but will have no impact on the content of the Mogul output file.

Acknowledgements

The CCDC gratefully acknowledges the following copyright works, which are used under licence in the CSD System.

Mesa © 1999-2001 Brian Paul
http://www.mesa3d.org/

Numerical Python © 1999, 2000, 2001. The Regents of the University of California.
http://www.pfdubois.com/numpy/

Pmw © 1997-2001 Telstra Corporation Limited, Australia
© 2000-2002 Really Good Software Pty Ltd, Australia
http://pmw.sourceforge.net/

pybsddb © 1999-2001, Digital Creations, Fredericksburg, VA, USA and Andrew Kuchling.
http://pybsddb.sourceforge.net/

PyOpenGL © 1997-1998, 2000-2001 by James Hugunin, Cambridge MA, USA, Thomas Schwaller, Munich, Germany and David Ascher, San Francisco CA, USA
http://pyopengl.sourceforge.net/

Python 2.2.1 © 2001, 2002 Python Software Foundation
http://www.python.org/

The Python Imaging Library (PIL)
© 1997-2002 by Secret Labs AB © 1995-2002 by Fredrik Lundh
http://www.pythonware.com/

Tcl/Tk © Regents of the University of California, Sun Microsystems, Inc., Scriptics Corporation, and other parties
http://www.scriptics.com/software/tcltk/

zlib © 1995-2002 Jean-loup Gailly and Mark Adler
http://www.gzip.org/zlib/

The CCDC acknowledge contributions made to Mogul by Richard Cooper and David Watkin, original author of CRYSTALS: *http://www.xtl.ox.ac.uk/crystals.html.*

Appendix A: Glossary

Analogues
Aromatic Bonds
Asymmetric Unit
Atomic Charge
Atomic Labels
Author(s)
Average Sigma (C-C)
Bioactivity
Bond Type Conventions
Cell Angles
Cell Lengths
Cell Volume
Chemical Formula
cif Format
Colour
Compound Name
con Format
Conformer
Connectivity Comment
Crystal Chemical Unit
Delocalised Double Bonds
Density
Disordered Structures
Experimental Notes
Fragment
Habit
Literature Reference
Melting Point
mol Format
mol2 Format
Molecule
pdb Format
Peptide Sequence
Phase Transitions
Pi-Bonds
Polymorph
Pressure
Radiation Source
Recrystallisation Solvent

Refcode (Entry ID)
Refinement Details
res Format
R-Factor
Sensitivity
Source
Space Group
Synonym
Temperature
Torsion Angles
Unit Cell Parameters
Z, Z'

Analogues

Where available, isostructural, isomorphous and isotypic analogues will be reported for hit structures. Details of isoelectronic, analogous or isomeric structures are not included.

Related Topics
- Displaying Information about Individual Structures .

Aromatic Bonds

An aromatic bond type is used for the ring bonds of benzenoid systems, 6-membered aromatic heterocycles, cyclopentadienyl rings, other ring systems that are pi-bonded to a metal ion, and 5-membered nitrogen heterocycles such as pyrazole and imidazole when they are sigma-coordinated through a ring nitrogen to a metal ion. However, 5-membered heterocycles such as thiophene, furan and non-metal-coordinated imidazole, etc., are normally coded with single and double bonds.

A few hundred CSD entries contain the 6-membered carbon ring with alternating single and double bonds. These structures are of 3 main types:
- Metal complexes where pi-bonding between the metal and the 6-membered carbon ring involves only 2 or 4 of the 6 carbon atoms.
- Phthalocyanines and other benzoporphines.
- Fullerenes

Related Topics
- Appendix B: Bond Type Conventions for Common Chemical Groups .
- Assignment of Unknown Bond Types and Missing Hydrogen Atoms .
- Auto-Editing Options .
- Bond Type Conventions .

- Manually Editing a Structure .

Asymmetric Unit

A crystal structure consists of a basic motif that is repeated in 3D space by the symmetry operators of the crystallographic space group. A crystallographer determines the coordinates of the atoms in this basic motif, called the *asymmetric unit*. It is the smallest part of a crystal structure from which the complete structure can be built using space group symmetry.

The asymmetric unit may consist of only one molecule or ion, part of a molecule, or several molecules that are not related by crystallographic symmetry. For example, consider structures of formula C_{12} H_{18} N_4 O_2:

- If the asymmetric unit contains one molecule, the crystallographer must determine the coordinates of 36 atoms.
- If the asymmetric unit contains two molecules, the crystallographer must determine the coordinates of 72 atoms.
- If the asymmetric unit is half a molecule, this implies that the molecule possesses symmetry coincident with a crystallographic symmetry element. For example, the molecule might possess a mirror plane, so that half of the atoms are related to the other half by symmetry. In this case, the crystallographer must determine the coordinates of only 18 atoms.

Related Topics

- Crystal Chemical Unit .

Atomic Charge

Where appropriate, atoms in CSD structures are assigned formal charges. In the CSD there is no concept of delocalised charge - charges are associated with specific atoms and must be integral. Atomic charges are ignored during a Mogul search. However, setting an atomic charge may affect the number of hydrogens that Mogul will attach to that atom when auto-addition of hydrogens is invoked.

Related Topics

- Assignment of Unknown Bond Types and Missing Hydrogen Atoms .
- Auto-Editing Options .
- Automatic Addition of Hydrogen Atoms
- Manually Editing a Structure .
- Required Molecular Information .
- Setting Atomic Charges .

Atomic Labels

Each atom in a structure has a label, which consists of the element symbol followed by a number and (sometimes) a prime. CSD atom labels are normally identical to, or closely related to, those used by the author. Atoms generated by symmetry (i.e. not belonging to the asymmetric unit) may have an extra letter at the end of their label that indicates which symmetry operation was used.

Related Topics

- Labelling Atoms in the 3D Visualiser .

Author(s)

Authors' names are stored exactly as given in the paper, with forenames abbreviated and stored as initials, e.g. F.Allen and F.H.Allen may both occur. Some author names may include dynastic tags, e.g. S.S.Simons Junior, A.J.Arduengo III.

Authors' names are stored in the CSD without umlauts, accents, etc. However, some journals express an umlaut by a following e; thus Müller, Sänger, etc., may sometimes appear as Mueller, Saenger, etc. in the CSD.

Chinese, Korean and Malaysian names are usually stored in full, e.g. Bing Bing Chang, Jung Mi Shin. Occasionally, however, Oriental names are recorded with initials, e.g. H.S.Kim. These variations are due to different journal conventions.

Names from non-Roman alphabets are not always transliterated in the same way in different papers, e.g. Belskii, Belsky, Belskij.

There is some inconsistency in the database in the handling of 2-letter initials, e.g. Yu.T.Struchkov might occasionally be stored as Y.T.Struchkov.

Related Topics

- Displaying Information about Individual Structures .

Average Sigma (C-C)

The average estimated standard deviation (e.s.d.) [standard uncertainty (s.u.)] of the carbon-carbon bond lengths in a structure provides a rough measure of precision and is quoted for most CSD entries. In certain entries, where no carbon-carbon e.s.d’s are available, then the value may be derived from C-N, C-O, N-N, N-O, O-O bond-length e.s.d.’s.

Related Topics

- Displaying Information about Individual Structures .

Bioactivity

For hit structures biological activity is reported if the author indicates that the compound, or a near-relative, is of biological interest. Where available. related information is also included e.g. not active; possible activity; biologically tested; derivative of or related to a compound with biological activity.

Related Topics

- Displaying Information about Individual Structures .

Bond Type Conventions

Mogul uses CSD bond-type conventions. Query structures must follow these conventions in order to retrieve all relevant hits. For example, a benzene ring in the CSD is coded using an aromatic bond type rather than alternate single and double bonds.

Related Topics

- Assignment of Unknown Bond Types and Missing Hydrogen Atoms .
- Auto-Editing Options .
- Appendix B: Bond Type Conventions for Common Chemical Groups .
- Bond Type Conventions .
- Manually Editing a Structure .

Cell Angles

For hit structures various unit cell data is recorded, including:

- alpha value of interaxial angle alpha (in degrees)
- beta value of interaxial angle beta (in degrees)
- gamma value of interaxial angle gamma (in degrees)

Related Topics

- Cell Lengths
- Cell Volume
- Displaying Information about Individual Structures .
- Unit Cell Parameters

Cell Lengths

For hit structures various unit cell data is recorded, including:

- a length of unit cell a-axis (in Angstroms)
- b length of unit cell b-axis (in Angstroms)
- c length of unit cell c-axis (in Angstroms

Related Topics

- Cell Angles
- Cell Volume
- Displaying Information about Individual Structures .
- Unit Cell Parameters

Cell Volume

For hit structures various unit cell data is recorded, including:

- v volume of unit cell (in cubic Angstroms)

Related Topics

- Cell Angles
- Cell Lengths
- Displaying Information about Individual Structures .
- Unit Cell Parameters

Chemical Formula

Since a crystal structure may contain more than one type of molecule, there are basically two sorts of formulae in the CSD:

- The formula of an individual molecule, for example $C_{10}\,H_{18}\,N_2\,Ni\,O_5$.
- The *sum formula*, i.e. the formula of all the different molecules in a structure added together. For example, the dihydrate $C_{10}\,H_{18}\,N_2\,Ni\,O_5,\,2(H_2O)$ would have a sum formula of $C_{10}\,H_{22}\,N_2\,Ni\,O_7$.

The letter *n,* which may sometimes be seen in formulae displayed in *View structures*, indicates that the compound is polymeric*,* for example $(C_3\,H_3\,O_6\,Sc)n$. The formula enclosed in the brackets is that of the monomer repeat unit.

Related Topics

- Displaying Information about Individual Structures .

cif Format

The Crystallographic Information File (CIF) format was developed as the standard crystallographic data exchange format (Hall, Allen and Brown, *Acta Cryst.*, **A47**, 655, 1991; *http://www.iucr.org/iucr-top/cif/home.html*).

Related Topics
- Input File Formats .

Colour

Where reported, the colour of the crystal at room temperature in daylight is given for hit structures.

Related Topics
- Displaying Information about Individual Structures .

Compound Name

Compound names in the CSD usually follow the rules of standard chemical nomenclature. Occasionally, a trivial, drug or trade name might be used.

A few conventions are used in constructing compound names:

- Bridging ligands in polymeric metal coordination complexes are identified by the bridging indicator *mu*, with the polymer identified by the prefix *catena*, for example *catena-((mu$_2$-2,5-dihydroxy-p-benzoquinonato)-zinc)*.
- Names of hydrates will contain the words *hemihydrate, monohydrate, dihydrate*, etc., otherwise, just *hydrate* if the multiplier is a non-integer value.
- If other solvents are present, the name will contain the word *solvate*; *clathrate* is used for solvates which are clathrated, as in host-guest compounds.
- Deuterated species will always contain the name characters *deuter*.

Related Topics
- Displaying Information about Individual Structures .

con Format

con is the format used for storing 2D chemical structures drawn in ConQuest. It can be output by using the **File ... Export QUEST Query...** option in the ConQuest Draw window.

Related Topics
- Input File Formats .

Conformer

When reported, stereodescriptors describing the shape of the whole molecule or structure are available for hit structures. Terms such as 1,2-alternative, cone, pinched cone, right-handed, helix, head-tail are used. Such text occurs frequently in the description of calixarenes, peptides and organic polymers, and may relate to configuration.

Related Topics

- Displaying Information about Individual Structures .

Connectivity Comment

As a result of CCDC validation, editorial comments relating to chemical connectivity may be included for some hit structures.

Related Topics

- Displaying Information about Individual Structures .

Crystal Chemical Unit

In some cases, molecular symmetry coincides with the symmetry elements of a crystallographic space group. For example, a molecule might have a mirror plane which coincides with a crystallographic mirror plane in the space group. In the crystal structure, the molecule has *exact m-symmetry*, the asymmetric unit is half a molecule, and the crystallographer only determines the coordinates of half of the atoms in the molecule.

In this situation, the atoms in the asymmetric unit plus the symmetry-generated atoms forming the other half of the molecule collectively form the crystal chemical unit.

Related Topics

- Asymmetric Unit .
- Displaying Crystallographic Unit-Cell Contents .

Delocalised Double Bonds

Delocalised double bonds are used in the CSD to describe the bonding in some delocalised structures. An example is metal-coordinated carboxylate. Rather than drawing one of the C-O bonds as single and one double (i.e. O-C=O), both bonds are assigned the *delocalised double* bond type (O--C--O). This is an attempt to reflect the fact that the C-O bonds are equivalent to each other.

The correct bond types must be specified in a Mogul query structure, otherwise relevant hit structures will not be found. Unfortunately, the CSD is often inconsistent in its use of delocalised double bonds. For example, about 50% of metal-bound acetylacetonato ligands are coded with the 6-membered ring having 4 single bonds and 2 double bonds. The other 50% have 2 single bonds and 4 delocalised double bonds.

Related Topics

- Appendix B: Bond Type Conventions for Common Chemical Groups .

- Assignment of Unknown Bond Types and Missing Hydrogen Atoms .
- Auto-Editing Options .
- Bond Type Conventions .
- Manually Editing a Structure

Density

This is the density of the crystal, as reported by the author or calculated from the reported chemical formula and unit cell data, using the relationship:

Density = (1.66 x formula weight x Z) / unit cell volume

where Z is the number of molecules in the unit cell.

Related Topics

- Displaying Information about Individual Structures .

Disordered Structures

Disordered structures display a lack of regularity. For example, each of the F atoms in the trifluoromethyl group -CF_3 might be randomly distributed between 2 sites. This means that the crystallographer will report two sets of coordinates for each F atom. In some cases, two alternative sites are occupied equally; in other cases, there is a major site and a minor site. Disorder can involve more than two sites and it can also involve a whole molecule.

Related Topics

- Displaying Information about Individual Structures .

Experimental Notes

When reported by the author, experimental details describing the crystal before diffraction are available for hit structures. Typical examples are: irradiated product, superconducting action, humidity, ground state, mesogen, dye, explosive, piezoelectric, triboluminescent, luminescent, monomerization study and thermal decomposition study.

Related Topics

- Displaying Information about Individual Structures .

Fragment

Fragment is the generic word used in Mogul for a bond, valence angle or torsion. The distribution of the length or angle of a particular fragment, as observed in CSD entries containing that fragment, can be determined by Mogul searching.

Related Topics

- Browsing the Chemical Structures of Hits .
- Chemical Coverage: The Mogul Library .
- Required Molecular Information .

Habit

When reported by the author, the crystal habit (i.e. the shape of the crystal, such as *needle* or *plate*) is stored for hit structures.

Related Topics

- Displaying Information about Individual Structures .

Literature Reference

Journal names in the CSD are normally the abbreviations adopted by the International Standards Organisation.

The page number is the starting page of the publication. Most journal pages are numeric but sometimes they contain letters, for example *L25, 123S*.

For about 20 journals, mostly Russian and Chinese, each issue starts at page 1. To provide an unambiguous reference, we store the issue number as well as the page number in the journal page field, separating them by a hyphen, for example *89-3* indicates page 89 of issue 3.

Since 1997, *Acta Cryst.,Sect.C (Cr.Str.Comm.)* has reported so-called *CIF-access papers*. These are represented in the Table of Contents of each issue by a brief abstract and diagram. The actual paper does not appear in the printed issue but can be retrieved from the IUCr archive using the printed journal data validation number. An example of such a number is IUC9900004. In this case the journal page in the CSD would contain 9900004.

J.Chem.Res. is published in two parts - the synopsis (S) and the miniprint (M). Since this journal has no volume number, the page number for part S is stored in the journal volume field and the page number for part M is stored in the journal page field.

The journal volume is normally numeric but may occasionally contain letters, for example *C471*. However, in other cases the alphabetic part of a volume number is absorbed into the journal name, for example *Acta Cryst., Section B, volume 47*.

For most journals, the volume takes a single value for each journal year. However, some journals have more than one volume per year and others have volume numbers that span a year change.

Some journals have no volume number, for example journals of the UK Royal Society of Chemistry, such as *J. Chem.Soc., Dalton Transactions*. Volume numbers are absent from PhD theses and private communications.

Related Topics
- Displaying Information about Individual Structures .

Melting Point

When reported by the authors, the melting point of hit structures is shown (either as a single point or a range, and either in Celsius or Kelvin depending on what the authors report).

Related Topics
- Displaying Information about Individual Structures .

mol Format

mol is a molecular file format produced by MDL Information Systems Inc., San Leandro, CA, USA.

Related Topics
- Input File Formats .

mol2 Format

mol2 is a molecular file format produced by Tripos Inc., 1699 South Hanley Road, St Louis, MO, 63144-2913 USA.

Related Topics
- Exporting Entries .
- Input File Formats .

Molecule

The word *molecule* in this document is used generically to refer to a molecule or ion. For example, sodium acetate monohydrate has 3 molecules: $C_2H_3O_2^-$, Na^+, H_2O.

pdb Format

pdb is the Protein Data Bank format for 3D structures.

Related Topics
- Exporting Entries .
- Input File Formats .

Peptide Sequence

The CSD covers peptides of up to 24 residues. Alpha-amino acids and modified alpha-amino acids (also zwitterionic and ionic) are represented by 3-letter codes and symbols; when linked by peptide bonds or non-peptide bonds into a peptide sequence the structure is represented by a peptide sequence of codes, symbols and links reading from the 'N' end of the molecule.

Peptide sequence can represent both cyclic and acyclic (linear) sequences and can contain undefined residues (UND) and branch-point symbols. Bicyclic (or multi- cyclic) peptides, pseudopeptides and retropeptides can also be represented.

Related Topics

- Displaying Information about Individual Structures .

Phase Transitions

When reported by the author, information regarding the temperature(s), pressure or other conditions of phase transition(s) are included for hit structures.

Related Topics

- Displaying Information about Individual Structures .

Pi-Bonds

Pi-bonds in the CSD refer to bonds between a metal and a ligand. For a cyclopentadienyl ring with no substituents or just acyclic substituents, a pi-bond is recorded between the metal atom and each of the ring C atoms. Similarly, butadiene would have 4 pi-bonds recorded to the metal atom.

The correct bond types must be specified in a Mogul query structure, otherwise relevant hit structures will not be found.

Related Topics.

- Appendix B: Bond Type Conventions for Common Chemical Groups .
- Assignment of Unknown Bond Types and Missing Hydrogen Atoms .
- Auto-Editing Options .
- Bond Type Conventions .
- Manually Editing a Structure

Polymorph

Polymorphism is the occurrence of two or more crystalline forms of the same substance. Hit structures known to be polymorphic contain comments which include the word polymorph (when

reported by the author), e.g. non-triboluminescent polymorph.

Related Topics
- Displaying Information about Individual Structures .

Pressure

When reported by the author, details of low or high pressure studies are recorded for hit structures.

Related Topics
- Displaying Information about Individual Structures .

Radiation Source

When reported by the author, data or refinement specifications are included to indicate:
- synchrotron radiation
- neutron
- neutron and x-ray radiation

References to electron radiation may be found for certain hit structures, though the decision to cease input of these studies was made in January 2001. Mo, Cu or wavelength data (to distinguish more than one study) are only very occasionally available.

Related Topics
- Displaying Information about Individual Structures .

Recrystallisation Solvent

When reported by the author, comments will be included for hit structures to indicate how crystals were prepared; also to indicate the source of natural products, e.g. from the melt, xtal hexane/diethyl ether, Flemingia fruticulose wall (Leguminosae).

Related Topics
- Displaying Information about Individual Structures .

Refcode (Entry ID)

Each CSD entry has a unique identifier known as a *refcode*. The refcode contains six letters, for example ABACOF. If more than one study of a compound is present in the CSD, the second and subsequent entries will have two numbers after the six letters, for example ABACOF01, ABACOF10, ABACOF03.

Deuterated forms of a compound have the same six-letter code as the non-deuterated form. Stereoisomers have different 6-letter codes.

Related Topics

- Displaying Information about Individual Structures
- Selecting Individual Structures for Viewing .

Refinement Details

When reported by the author, information concerning the refinement can be available for hit structures eg. *refinement in centrosymmetric space group*. Other terms recorded include: multipole, high-angle, rigid body, final, 2, IV refinement, isotropic, anisotropic, Frenkel, model 2, full data, kappa, rigid body, high-order.

Related Topics

- Displaying Information about Individual Structures .

res Format

Crystal structure file format used by the program SHELX (*http://shelx.uni-ac.gwdg.de/SHELX/*).

Related Topics

- Input File Formats .

R-Factor

The crystallographic R-factor is the traditional figure of merit for crystal structures and provides a measure of how well the refined structure agrees with the experimental model. Authors often also report a weighted R-factor, wR. The value stored in the CSD is the lower of R and wR.

A rough guide to the quality of structure determinations is:

R-factor	*Quality*
0.01 - 0.03	Exceptional
0.03 - 0.04	Very high
0.04 - 0.05	High
0.05 - 0.07	Good
0.07 - 0.09	Average
0.09 - 0.10	Fair
0.10 - 0.15	Poor

R-factor	*Quality*
> 0.15	Bad

Related Topics

- Displaying Information about Individual Structures .
- Filtering Hits .

RMSD

Root Mean Standard Deviation calculation is a standard method of assessing how dissimilar two geometric arrangements of atoms are in two structures, where a one-to-one correspondance exists between the two sets of atoms. The same calcualtion can be applied to geometric features (e.g. ring torsions), if the same correpondance can be applied. The distance between each atom pair (or, for ring torsions, difference between corresponding ring torsions) is squared, these are summed and averaged, and then the square root taken to calculate the RMSD of one set against the other. RMSD's are always greater than zero.

Sensitivity

When reported by the author, properties of the crystal will be described for hit structures. Typically, comments are included for structures which are:
air- and moisture-sensitive
hygroscopic
efflorescent
deliquescent
heat-sensitive (needs to be stored below room temperature)
oxygen-sensitive
light-sensitive
photo-sensitive
pyrophoric

Related Topics

- Displaying Information about Individual Structures .

Source

When reported by the author, information relating to the natural-product source of the chemical compound or its immediate parent compound will be available for hit structures.

Related Topics

- Displaying Information about Individual Structures .

Space Group

There are 230 possible arrangements of symmetry elements in the solid state. They are called *space groups* (see *International Tables for Crystallography,* Volume A, Kluwer Academic Publishers, 1983). Any crystal must belong to one (and only one) space group.

The space groups are numbered from 1 to 230 and each is represented by a space group symbol; for example, space group number 19 has the symbol $P2_12_12_1$.

For certain space groups, it is possible to choose the unit cell axes and/or origin in more than one way (*alternative settings*). This means that a given space group number can correspond to several space group symbols. For example, space group number 25 corresponds to Pmm2, P2mm and Pm2m.

The following trigonal space groups can be described with respect to hexagonal or rhombohedral axes:

146	R3
148	R-3
155	R32
160	R3m
161	R3c
166	R-3m
167	R-3c

In the CSD, the rhombohedral setting is identified by appending the letter *r* to the space group symbol, for example R-3cr.

In certain situations there is an unresolved ambiguity in the space group determination. For example, if the crystallographer cannot decide between Pmaa and P2aa then the space group is represented by the so-called aspect symbol, in this case P*aa. There are 127 aspect symbols having *CSD space group numbers* in the range 502-765.

Related Topics

- Displaying Information about Individual Structures .

Synonym

When reported by the author, any appropriate synonym(s) for the compound name will be recorded.

Related Topics

- Compound Name
- Displaying Information about Individual Structures .

Temperature

A comment is included if a hit structure has been determined at a temperature other than room temperature. Any structure determined in the range 283-303K is considered to be a room-temperature structure.

Related Topics

- Displaying Information about Individual Structures .

Torsion Angles

Torsion angles are used to describe conformations around rotatable bonds. The torsion angle between 4 atoms A-B-C-D is the angle by which the vector A-B must be rotated in order to eclipse the vector C-D when viewed along the vector B-C. Crystallographers usually express torsion angles in the range -180 to +180 degrees. In Mogul, only the absolute values of torsion angles are used in histogram displays.

Related Topics

- Searching for an Individual Bond Length, Valence Angle, Torsion Angle or Ring .
- Viewing a Histogram and Selecting Hits .

Unit Cell Parameters

The unit cell is the basic building block of a crystal, repeated infinitely in three dimensions. It is characterised by:

- three vectors (*a, b, c*) that form the edges of a parallelepiped;
- the angles between the vectors (*alpha*, the angle between *b* and *c*; *beta*, the angle between *a* and *c*; *gamma*, the angle between *a* and *b*).

Depending on the crystal system there are sometimes restrictions on the values that unit cell parameters can take:

Triclinic/Anorthic	no restrictions

Monoclinic		alpha = gamma = 90
Orthorhombic		alpha = beta = gamma = 90
Tetragonal	a = b	alpha = beta = gamma = 90
Trigonal/Hexagonal	a = b	alpha = beta = 90, gamma = 120
Rhombohedral	a = b = c	alpha = beta = gamma
Cubic	a = b = c	alpha = beta = gamma = 90

In the above table, alpha=gamma=90 for monoclinic corresponds to the *b-axis unique* setting. Two other settings are possible:

a-axis unique: beta = gamma = 90
c-axis unique: alpha = beta = 90

Related Topics
- Displaying Information about Individual Structures .

Z, Z'

Z is the number of molecules in the crystallographic unit cell.
Z' is the number of molecules in the asymmetric unit.

Note: In crystal structures of solvates, ion pairs, clathrates and host-guest complexes, the word molecule should be taken to include the entire formula unit, e.g. both the metal-containing moiety and the water molecule in a compound of formula C_{10} H_{10} Fe_1 N_1 O_2, H_2O.

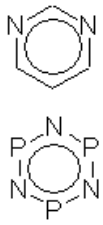
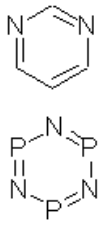
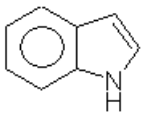
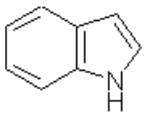
Related Topics:
- **Asymmetric Unit**
- **Crystal Chemical Unit**
- Displaying Information about Individual Structures .
- **Molecule**

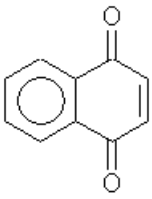
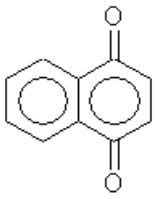
Appendix B: Bond Type Conventions for Common Chemical Groups

When searching Mogul, it is important that query structures follow the same bond-type conventions as the CSD itself. Very often, the **Auto-Edit** options in the Mogul *Build query* pane can be used to ensure that this is the case, but they occasionally fail. If you run a search that finds suspiciously few hits, it is worth experimenting with other possible bonding representations. These tables contain some guidelines and examples to help you. The CSD itself is not entirely consistent, so there is often not an absolute "right" or "wrong" answer. Most of the difficulty arises with the following systems:

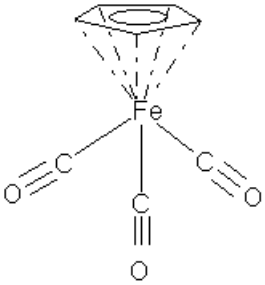
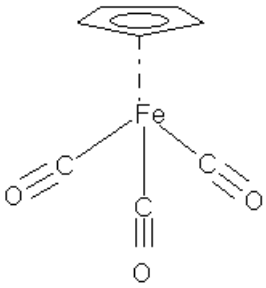
Aromatic systems
Pi-Bonded Metal Complexes
Difficult Functional Groups and Ions

Aromatic systems

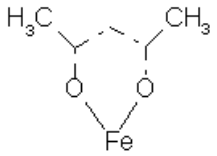
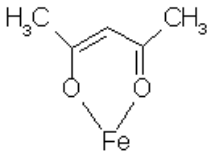
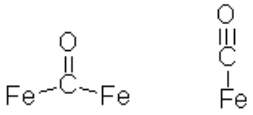
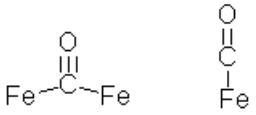
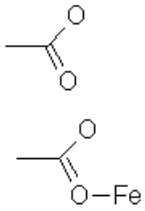
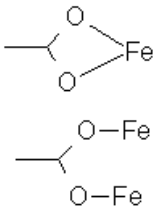
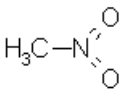
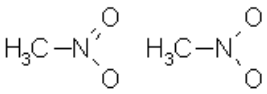
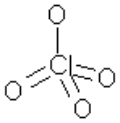
Group	Guidelines	Correct Example	Incorrect Example
Unfused aromatic 6-membered rings.	**Use the aromatic bond type, not alternate double and single bonds.**		
Aromatic 5-membered nitrogen heterocycles such as imidazole, pyrazole, when one of the ring nitrogens is sigma-bonded to metal; but excluding pyrrole rings of porphyrins	**Use the aromatic bond type.**		
Unfused aromatic 5-membered rings, except for rings that are pi-bonded to metal and metal-bound nitrogen heterocycles (see above).	**Use the appropriate combination of double and single bonds.**		
Fused aromatic ring systems.	**Use the aromatic bond type for all bonds in 6-membered rings; use single or double bonds as appropriate for any remaining bonds in 5-membered rings.**		

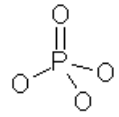
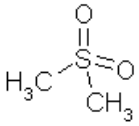
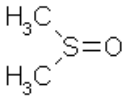
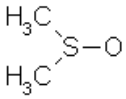
Group	Guidelines	Correct Example	Incorrect Example
Cyclopentadienyl rings and all other aromatic rings, of any size, that are pi-bonded to metal.	**Use aromatic bond type for all bonds.**		
Carbonyl-containing conjugated ring systems.	**Use double and single bonds as appropriate for the rings bearing the carbonyls; follow the conventions above for the other rings.**		
Porphyrins and related systems. Fullerenes.	**Use appropriate combination of double and single bonds.**		

Pi-Bonded Metal Complexes

Group	Guidelines	Correct Example	Incorrect Example
Pi-bonded metal complexes.	**Ensure that the metal atom forms a pi bond to every atom to which it is eta-connected, e.g. all 5 atoms of a cyclopentadienyl ring.**		

Difficult Functional Groups and Ions

Group	Guideline	Correct Example	Incorrect Example
Acetylacetonato ion, coordinated to metal.	**Use delocalised bond type for the carbon-oxygen and carbon-carbon bonds is the conjugated system.**		
Carbonyl.	**Use a double C=O bond if the group is bridging metal atoms, but use a triple bond if it is bonded to only one metal atom.**		
Carboxylate ion, uncoordinated or coordinated via only one of the oxygen atoms. Or thio equivalent.	**Use one single C-O bond and one double C=O bond.**		
Carboxylate ion, bidentate to one or two metals. Or thio equivalent.	**Use the delocalised bond type for both carbon-oxygen bonds.**		
Nitro and nitrate.	**Use two double N=O bonds (an uncoordinated nitrate ion would have two double bonds and one single).**		
Perchlorate ion.	**Use three double bonds and one single bond.**		

Group	Guideline	Correct Example	Incorrect Example
Phosphate, phosphonate, phosphinate ions.	**Use one double bond and three single.**		
Sulfone and sulfonamide.	**Use two double S=O bonds.**		
Sulfoxide.	**Use a double S=O bond.**		

Appendix C: Information Available for Individual Structures

Information such as the literature reference, chemical name, etc. may be available for individual hit structures; this information can be viewed by hitting the **Information** button in the *View structures* screen (see Displaying Information about Individual Structures).

The following data items are available:

Analogues	**Author(s)**	**Average Sigma (C-C)**	**Bioactivity**
Cell Angles	**Cell Lengths**	**Cell Volume**	**Chemical Formula**
Colour	**Compound Name**	**Conformer**	**Connectivity Comment**
Density (author)	**Density (CCDC)**	**Disordered Structures**	**Experimental Notes**
Habit	**Literature Reference**	**Melting Point**	**Peptide Sequence**
Phase Transitions	**Polymorph**	**Pressure**	**R-Factor**
Radiation Source	**Recrystallisation Solvent**	**Refcode (Entry ID)**	**Refinement Details**
Sensitivity	**Source**	**Space Group**	**Synonym**
Temperature	**Z, Z'**		

Note: Information corresponding to a specific data item will only be present when reported by the author in the original publication or deposited as supplementary data.

Appendix D: Tutorials

Tutorial 1 Determining Conformational Preferences: Performing a Torsion Angle Search
Tutorial 2 Finding More Hits
Tutorial 3 Validating Molecular Dimensions: Performing an All Fragments Search
Tutorial 4 Analysing the structure of a protein bound ligand: Performing a Ring Search

Tutorial 1 Determining Conformational Preferences: Performing a Torsion Angle Search

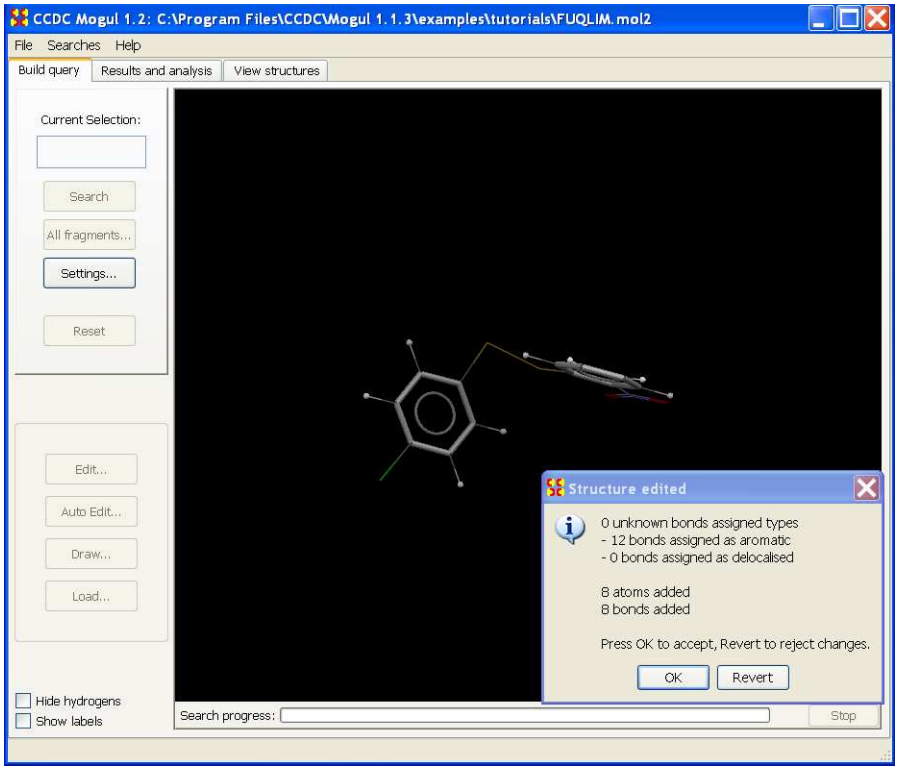
The Example

- The generation of torsion angle distributions to determine conformational preferences about single rotatable bonds is one of the most common uses of the CSD, particularly for molecular modelling applications. Indeed, CSD-based torsion angle data are used in programs that generate low-energy conformations and in protein-ligand docking applications (where CSD torsional distributions are used to bias docking solutions towards favourable ligand geometries).
- This tutorial illustrates how Mogul can be used to rapidly determine the disulfide torsion angle preferences of 2-nitro-4'-chlorodiphenyl disulfide by inspecting similar structures in the CSD.

Menu Commands Required

1. Import the query structure.
   - A query molecule can be submitted to Mogul directly (i.e. it is not necessary to construct a query manually as in ConQuest). A query structure will typically consist of a single molecule, or multiple molecules or ions. Both 2D and 3D structures can be used and a range of file formats are accepted (see Importing a Query Structure).
     Note: 2D queries can also be sketched using the Mogul *Draw* window.
   - Click on the **Load** button in the *Build query* pane. In the resulting *Load molecule* dialogue box, select `FUQLIM.mol2` from `<SOFTWARE_INSTALLDIR>\examples\tutorials\` and hit **Open.**

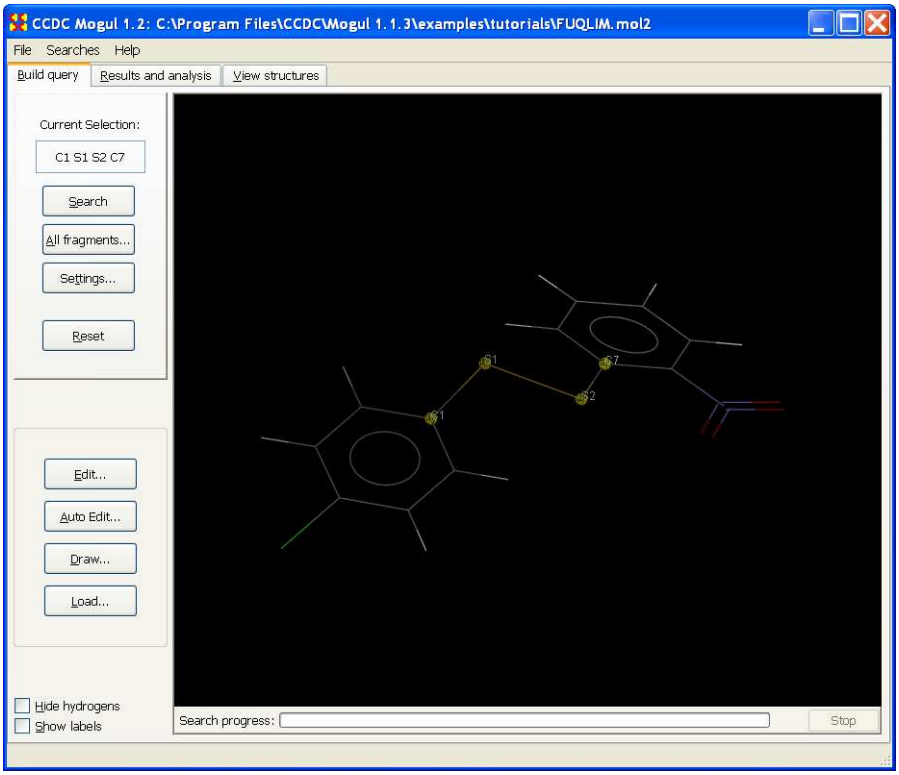
2. Assignment and standardisation of bond types and addition of hydrogens.
   - The results of a Mogul search will be erroneous if the query does not have correct bond types or is missing hydrogen atoms. Therefore, on loading a structure, Mogul will automatically:
     - Guess any bond types that are not specified in the input file.
     - Standardise all bond types according to CSD conventions (i.e. aromatic and delocalised bond type will be used where appropriate).
     - Add any missing hydrogen atoms.
   - Any changes made to an input structure are summarised in the *Structure edited* dialogue box:



- For the 2-nitro-4'-chlorodiphenyl disulfide input structure bonds in the two 6-membered aromatic rings have been converted to the CSD *aromatic* bond type and eight hydrogen atoms not present in the input file have been added.
- Hit **OK** within the *Structure edited* dialogue box to accept these changes.
  Note: there is no guarantee that bond-type deduction or hydrogen addition will be completely correct, hence it is also possible to edit structures manually (see Editing a Structure).

3. Define the search fragment (i.e. select the disulfide torsion angle).

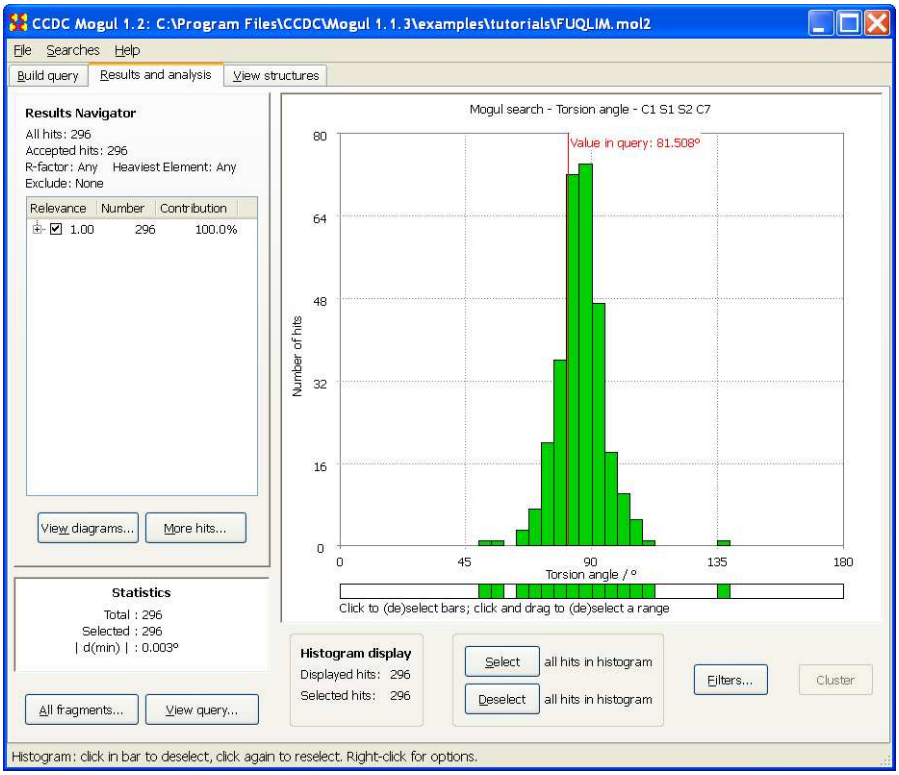
- Select the four atoms that are needed to define the C-S-S-C torsion angle by clicking on them with the left-hand mouse button (if you make a mistake hit **Reset** to clear the current selection).
- Selected atoms will be highlighted within the query. The atoms selected will also be listed under **Current Selection** on the left of the *Build query* pane:



- A search substructure will be generated automatically based on the specified search fragment (i.e. the selected torsion angle). The substructure will extend outwards from the search fragment by two bonds in all directions (i.e. the chemical environments of all atoms bonded directly to the C-S-S-C atoms are considered) (see Required Molecular Information).
- Hit **Search** to run the search.

4. Viewing and analysing the results.

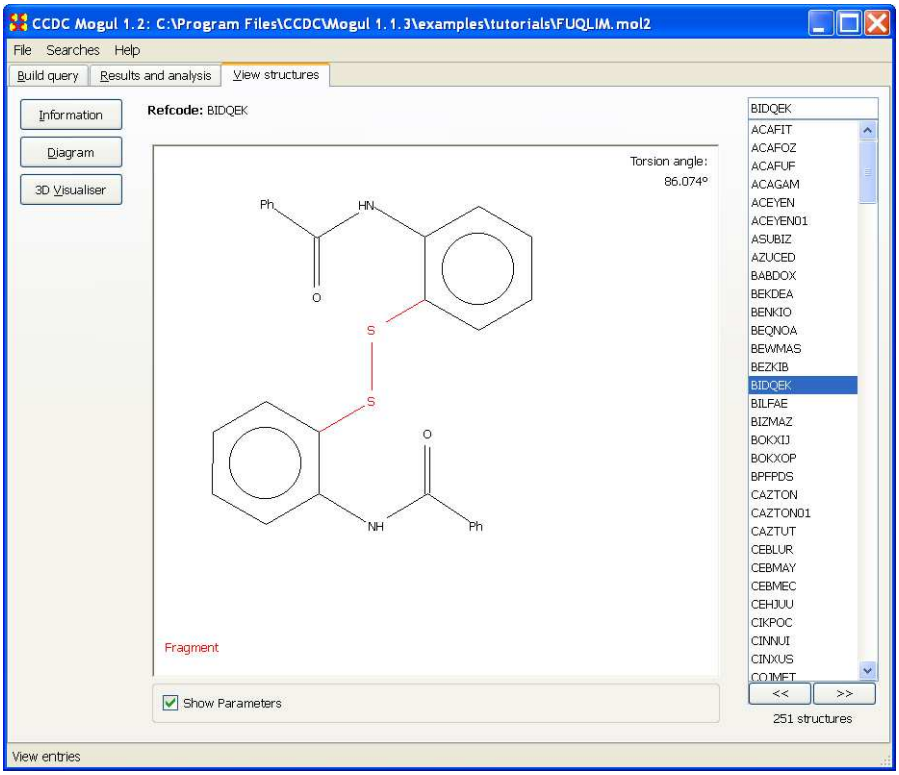
- The resulting histogram shows the torsion angle distribution calculated from the CSD entries that match the input query (i.e. those CSD entries that contain the same disulfide moiety as the query molecule):



- The distribution shows the preferred disulfide torsion angle to be 90° (only the absolute values of torsion angles are plotted in the histogram).

5. Viewing hit structures.

- By default, all the CSD entries contributing to the histogram are listed and can be viewed by clicking on the *View structures* tab. However, this list can be restricted to just those CSD entries that contribute to one or more chosen bins.
- By default all bins are selected. Deselect all bins in the histogram by clicking on the **Deselect** button. Then, select some of the individual bins within the histogram by clicking on them with the left mouse button.
- Summary statistics for selected hits in the histogram are displayed on the left hand-side of the *Results and analysis* pane. Notice that these statistics are automatically updated to reflect the bins currently selected.
- Click on the *View structures* tab and inspect some of the CSD entries that contribute to the bins currently selected in the histogram. To view a particular structure found by a Mogul search click on its refcode (CSD entry identifier) in the list on the right-hand side of the *View structures* pane:



- For each hit structure information such as the literature reference, chemical name, etc. may be available. A 2D chemical diagram and the 3D structure can also be viewed (see Viewing Hit Structures).

Tutorial 2 Finding More Hits

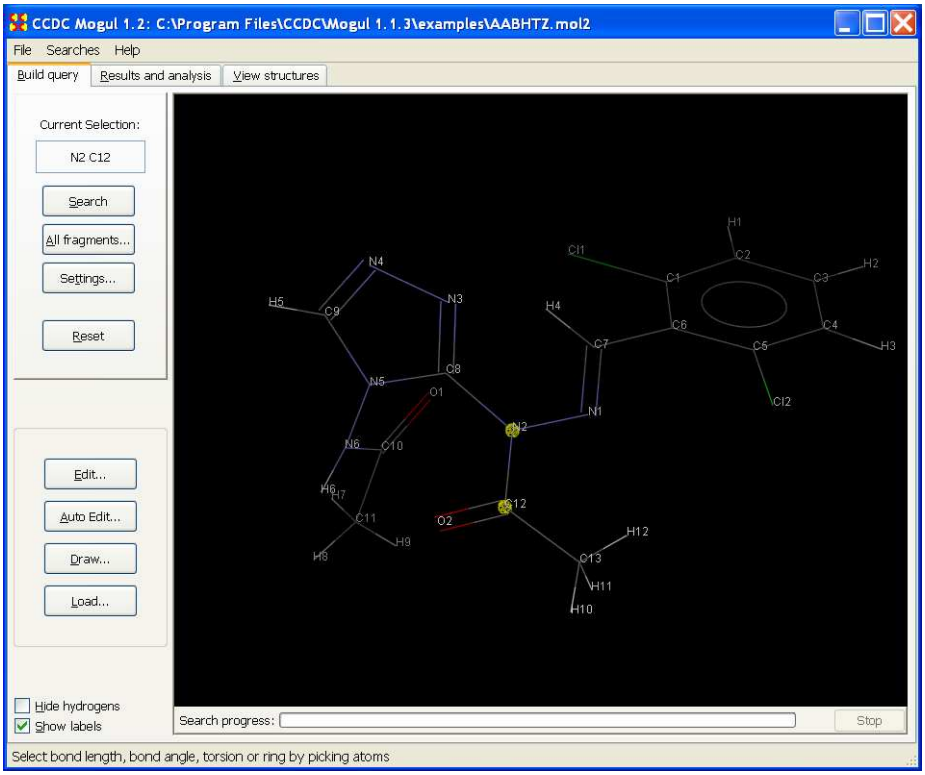
The Example

- When searching on a particular geometric feature, the number of hit fragments found that are structurally identical to the query fragment may not be enough. In such cases, Mogul will by default look for fragments that, while not identical to the query, are sufficiently structurally related as to be relevant.
- If the number of hits found are still not sufficient, it is possible to request additional hits.
- This tutorial demonstrates how to find additional hits for the C12-N2 bond fragment in 4-acetoamido-3-(1-acetyl-2-(2,6-dichlorobenzylidene)hydrazine)-1,2,4-triazole (CSD refcode AABHTZ).

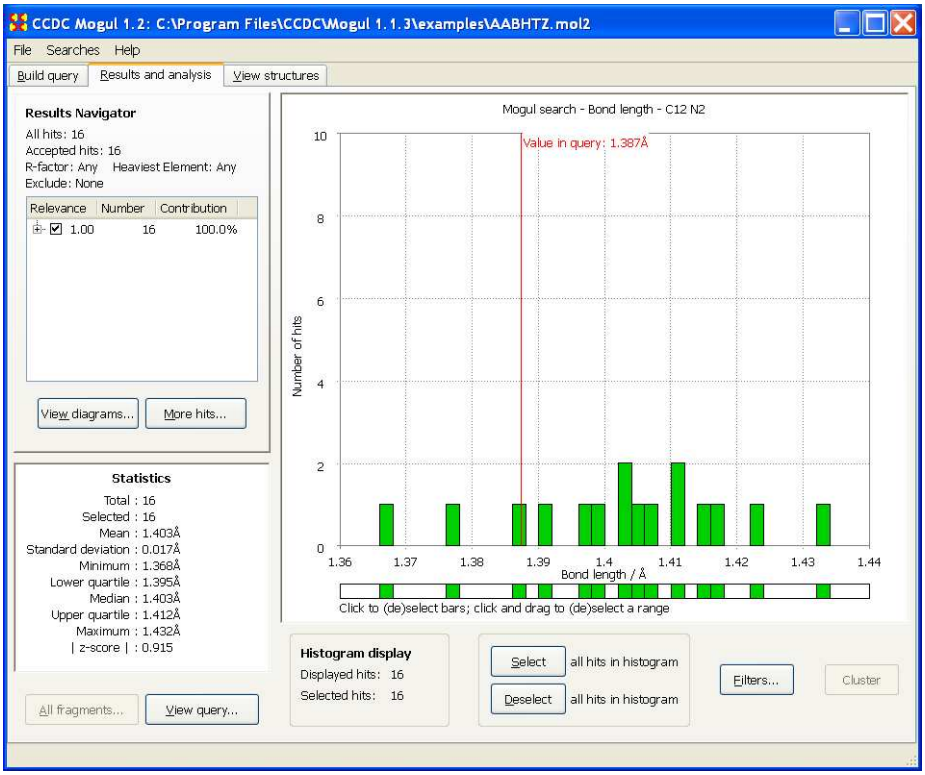
Menu Commands Required

1. Import the query structure.
   - Click on the **Load** button in the *Build query* pane. In the resulting *Load molecule* dialogue box, select `AABHTZ.mol2` from `<SOFTWARE_INSTALLDIR>\examples\tutorials\` and hit **Open.**

2. Define the bond length and run the search.
   - Select the two atoms required to define the C12-N2 bond length (you can display the atom labels for the query structure by enabling the **Show labels** check-box in the bottom left corner of the *Build query* pane).
   - Hit **Search** to start the search.



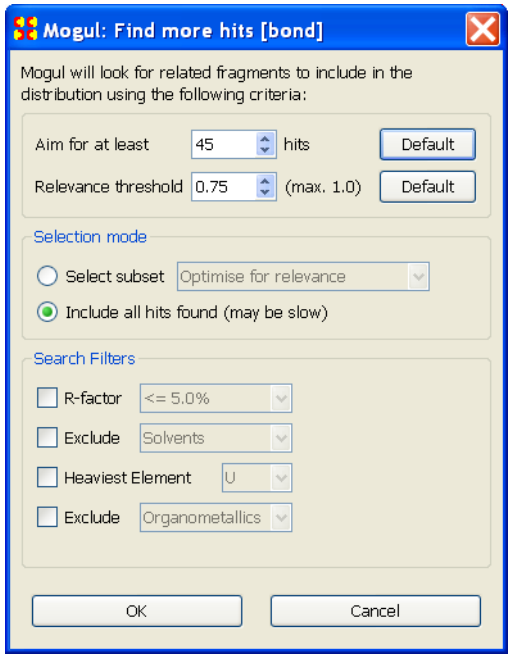
3. Inspect the results.
   - The resulting histogram contains relatively few hits and although these results are meaningful, the total number of observations in the distribution is possibly insufficient.



- This problem is overcome by requesting additional hits. *Generalised* searching, is carried out by default, and allows Mogul to look for fragments that, while not identical to the query, are sufficiently closely related as to be relevant. These fragments may then be incorporated into a *generalised* distribution, which therefore consists of observations from fragments that are similar to the query fragment as well as fragments that are identical to the query (.
- By default, a generalised search will be carried out if (for bond and angle fragments) less than 15 exact observations are found, and the generalised search will aim to find at least 15 observations. If the number of generalised hits found is still not sufficient, it is possible to request additional hits. Here we have rerieved 16 identical hits to the search query

4. Finding more hits

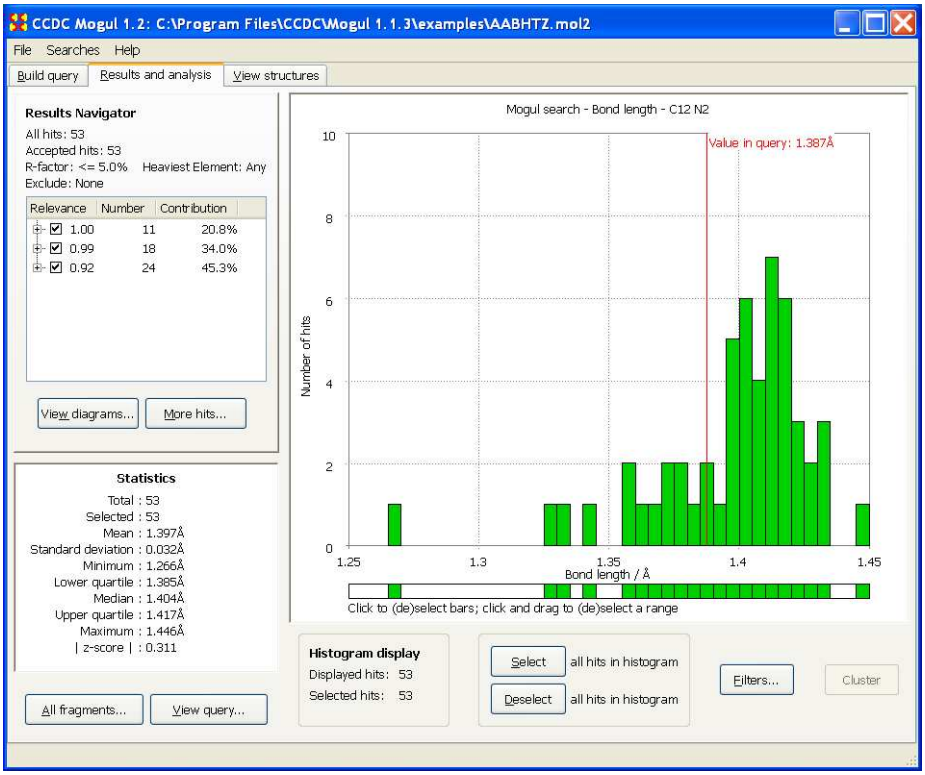
- Click on the **More hits...** button in the *Results and analysis* pane. The resulting *Mogul: Find more hits* window contains a number of settings that are used to control the *generalised* search:



- Specify that the *generalised* distribution should contain at least 45 observations by entering 45 in the **Aim for at least** *x* **hits** box (this number is the target number the search will aim for).
- Edit the **Relevance threshold** box to read 0.85, so that we leave out the lowest relevance hits.
- The *selection mode* provides control over the speed of the *generalised* search, the relevance of the hits found and the size of the resulting distribution. Select **Optimise for relevance** from the **Select subset** drop down menu, this will instruct Mogul to find only the most relevant fragments but may result in slow search speeds (see Controlling the Speed and Quality of Generalised Searches).
- Restrict the search to only those structures with R-factors less than 5% by selecting **< 5%** from the **R-factor** drop-down menu.
- Hit **OK** to run the *generalised* search.

  Note: For further information on *generalised* search settings, see **Finding More Hits**.

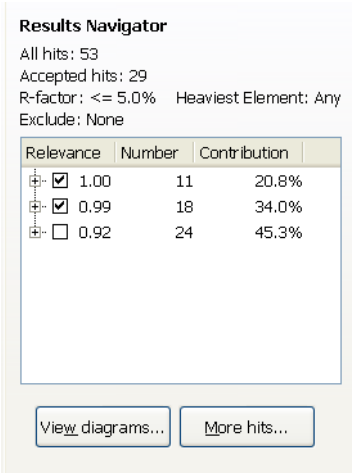
5. Viewing the *generalised* search results.
   - The size of the resulting *generalised* distribution should be close to the value specified in the **Aim for at least** *x* **hits** entry box (i.e. 45).



- For *generalised* searches Mogul determines how closely related a particular fragment is to the query fragment by calculating its "relevance" (a number between 0 and 1; higher values imply closer similarity to the query).
- Fragments contributing to a *generalised* distribution are listed in the **Results Navigator** located on the left of the *Results and analysis* pane.
- By default, fragments are listed in descending order of relevance (so any exact matches will come first, with a relevance of 1.0). The number of fragments with a given relevance is shown, together with their percentage contribution to the total distribution.

6. Browsing hit structures.

- Groups of contributing entries can be switched on and off using their corresponding check-box within the **Results Navigator**. The histogram and summary statistics are automatically updated to reflect the fragments currently selected.
- Using the appropriate check-boxes, deselect all entries with a relevance greater than 0.95 (so that only fragments least similar to the query are displayed in the histrogram):



- Click on the **View diagrams...** button to the left of the *Results and analysis* pane to view the 2D chemical structures of those hits currently selected.
- Use the << and >> buttons to browse the hit structures. At this relevance value are you satisfied that the hits are similar enough to the query fragment to be useful?

7. Specifying and saving search settings.

- *Generalised* searching is performed by default (in situations where the number of *exact* hits is insufficient).
- Click on the **Settings...** button in the *Build query* pane. In the resulting *Mogul search settings* pop-up, select the *General* tab and see that the **Generalised** search type check-box is selected.
- The criteria used to control generalised searches can then be set independently for bond, angle and torsion searches.
- Select the *Bond* tab in the *Mogul search settings* dialogue box. Ensure that the bond fragment search criteria are set-up such that:
  - A *generalised* search will be performed if fewer than 15 *exact* matches are found.
  - The target number of observations in a *generalised* distribution is at least 15.
  - Only fragments with a relevance of at least 0.75 will be included in a *generalised* distribution.
  - Only the most relevant fragments will be included unless this causes the search to take a long time.
- Click on the *General* tab of the *Mogul search settings* dialogue box and save your search settings by enabling the **Save settings on exit from Mogul** check-box..

Tutorial 3 Validating Molecular Dimensions: Performing an All Fragments Search

The Example

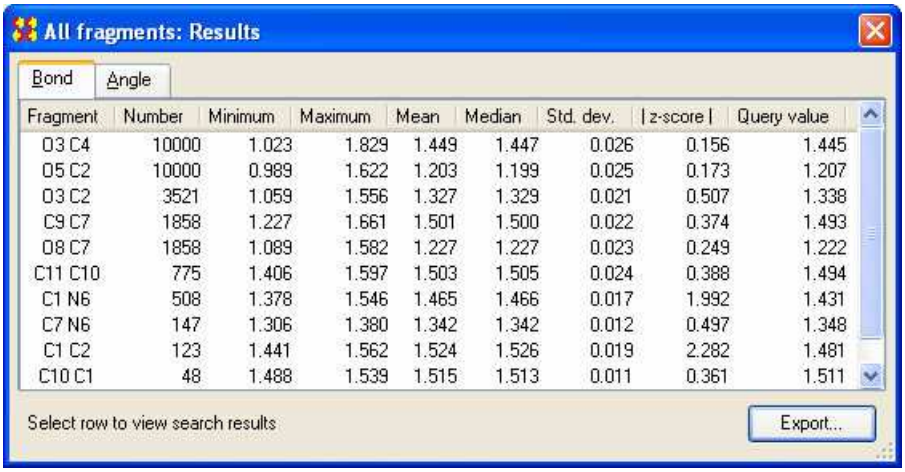
- Comparing the dimensions of a newly determined small-molecule crystal structure with the bond lengths and angles of similar structures in the CSD is extremely useful both as a check against refinement errors and to highlight unusual geometric features.
- This tutorial demonstrates how to search on all bond and angle fragments in a query molecule and shows how unusual or even suspect geometric features can be readily identified.

Menu Commands Required

1. Import the query structure.
   - Click on the **Load** button in the *Build query* pane. In the resulting *Load molecule* dialogue box, select `cyclopropyl.mol2` from `<SOFTWARE_INSTALLDIR>\examples\tutorials\` and hit **Open.**

2. Search on all bond lengths and angles.
   - An *All fragments* search will allow you to search for all valid bond lengths, valence angles, torsion angles, and/or unfused rings within your query molecule.
   - Click on the **All fragments...** button on the left of the *Build query* screen. In the resulting *Search for all fragments* pop-up, disable the **All torsion fragments** check-box and the **All ring fragments** check-box, and ensure that both the **All bond fragments** and **All angle fragments** check-boxes are selected.
   - Hit **Search** to run the search.

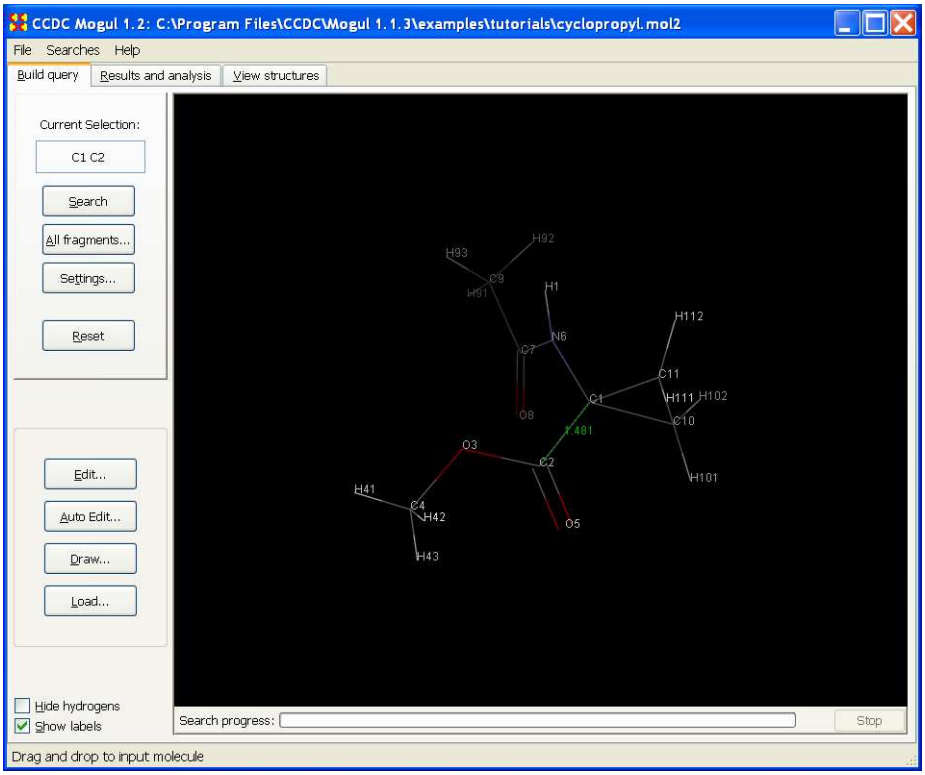
3. Viewing the search results.
   - The results from an *All-fragments* search are displayed (in spreadsheet format) in a separate *All-fragments search* window.



- Results for either bond length or valence angle fragments can be viewed by clicking on the appropriate tab at the top of the *All-fragments search* window.
- Click on the *Bond* tab. Each bond fragment in the query structure is listed in the spreadsheet together with the summary statistics for the corresponding Mogul distribution.

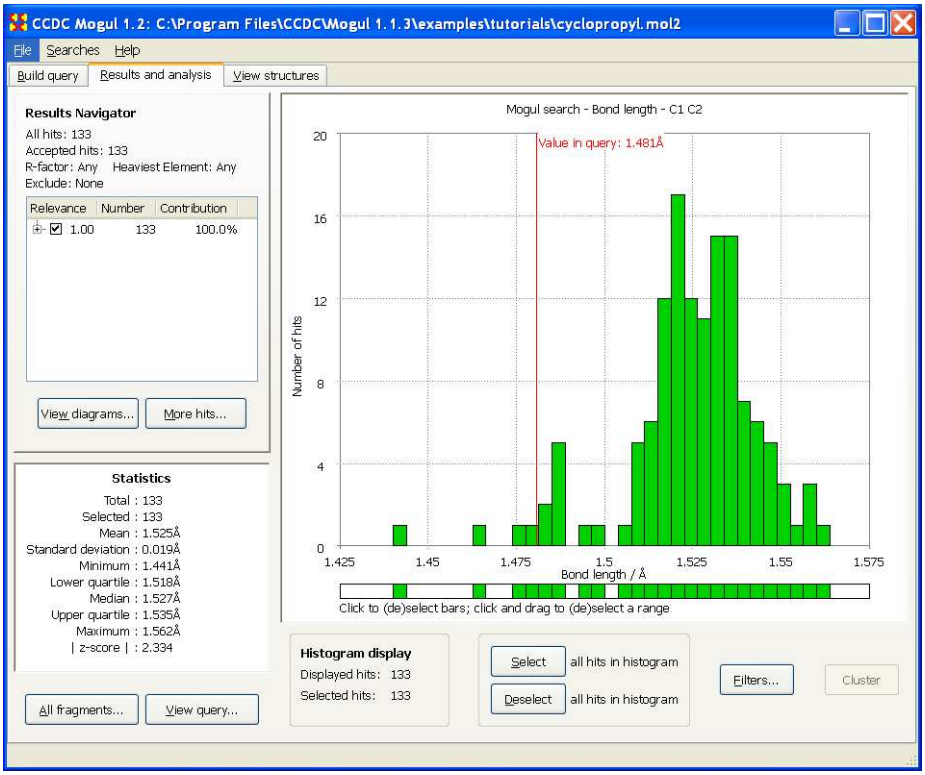
4. Identifying unusual geometric features.

- For each bond fragment in the query structure statistics are given in the *All-fragments search* window, these include: number of observations, minimum, maximum, mean, median, standard deviation, value in query and z-score.
- z-score is the absolute difference between observed and mean values of a geometric parameter divided by the standard deviation of the Mogul distribution. Therefore, a high z-score (e.g. >2.0) may indicate an unusual or even suspect geometry within your query.
- The rows of a spreadsheet can be sorted according to the values in any of the columns. Click on the **z-score** column header button to sort the rows by z-score.
- Notice that the C1-C2 bond length has a high z-score value (2.334). To investigate this potentially suspect bond length further, display the search results for the C1-C2 bond fragment by clicking on the corresponding row in the spreadsheet.
- In the main *Mogul* window click on the *Build query* tab. The C1-C2 bond fragment is directly attached to a cyclopropyl ring and is highlighted in the query structure.

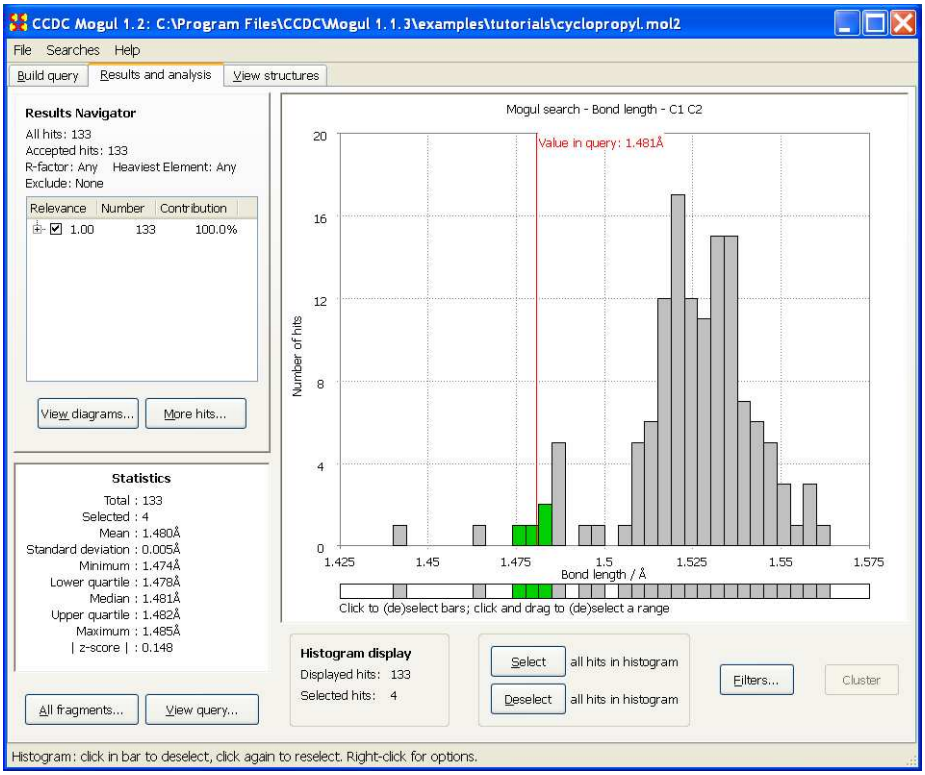


5. Analysing the results

- Click on the *Results and analysis* tab in the main *Mogul* window. The value of the C1-C2 bond length in the query is superimposed in red on the histogram. This allows for easy comparison with the geometric results obtained from Mogul.



- When compared to similar structures in the CSD the C1-C2 bond in the query structure appears to be unusually short (1.481 Å) and falls outside the main Mogul distribution.
- However, there are a small number of observations in the histogram with a similar value to that of the query structure.
- In order to inspect just these structures deselect all hits in the histogram by clicking on the **Deselect** button, then highlight the three histogram bins located around 1.48 Å using the horizontal bar located directly under the histogram:



- Click on the *View structures* tab and inspect the CSD entries that contribute to the selected bins.
- Notice that for three of the four hit structures the search fragment is also attached directly to a cyclopropyl ring. Therefore, it might be reasonable to assume that the shortening of the C1-C2 bond is a consequence of the cyclopropyl group and representative of this type of structural motif (i.e. the C1-C2 bond length in the query structure is in fact correct).
- In order to confirm this check some of the hit structures in the more populated region of the distribution and satisfy yourself that these do not contain a cyclopropyl group.

Here ends the tutorial

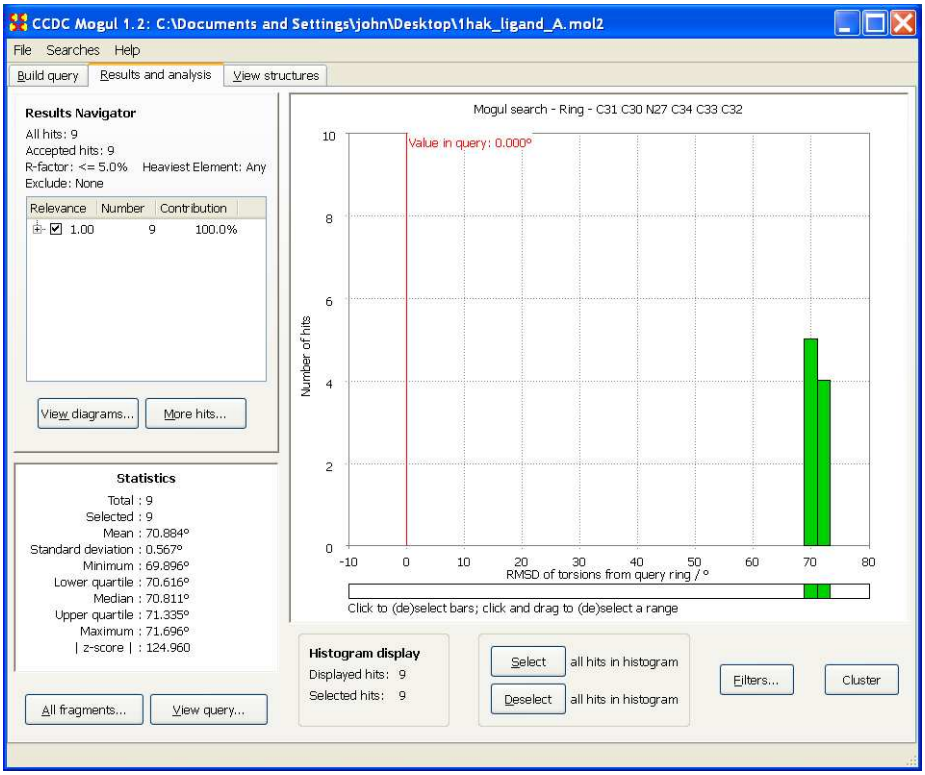
Tutorial 4 Analysing the structure of a protein bound ligand: Performing a Ring Search

The Example

- The refinement of protein/ligand structures from X-ray diffraction patterns is clearly a harder problem than refinement from small molecule derived diffraction patterns. The resolution of these patterns is lower as a rule and consequently it is very rarely possible to get electron density resolved to the atomic level. It is therefore necessary to assume standard bond lengths and bond angles in a protein refinement and use these to help the refinement process generate reasonable models. Quite frequently however ligand models are generated which are far from being low energy structures. Occasionally this might be because the protein is straining the ligand. More frequently it is because alternative ligand models that might have fitted the electron density equally well, but with lower strain energy, have not been investigated
- Mogul can be used to examine structures of ligands bound to proteins against similar chemistry within the CSD. An assessment can be made as to whether a ligand structure of unusual geometry is correct and strained by the protein, or alternatively, is unlikely to be correct. A thorough assessment should normally look not only at the ligand structure but also the resolution of the complex, the fit of the ligand structure to the electron density of the complex by visual inspection, and possibly an analysis of the quality and nature of the interactions made by the ligand to the protein.
- This tutorial illustrates how Mogul can be used to assess two models of a ligand structure from the protein data bank entry *1hak*.

Menu Commands Required

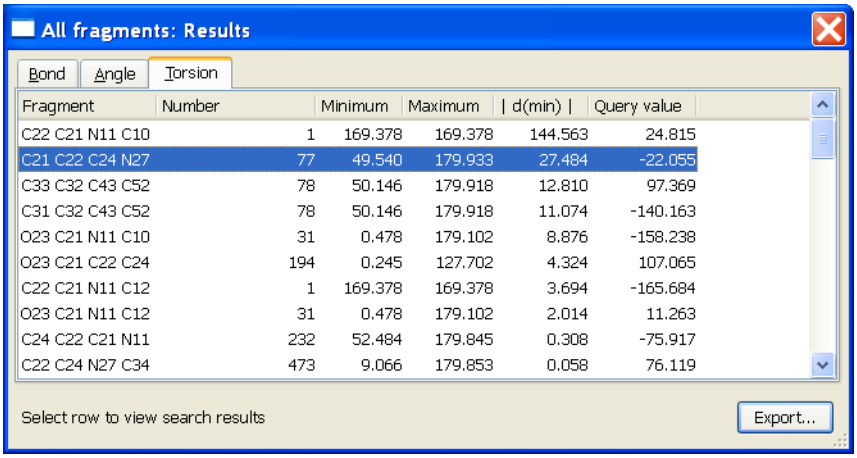
1. Import the query structure.
   - Click on the **Load** button in the *Build query* pane. In the resulting *Load molecule* dialogue box, select `1hak_ligand_A.mol2` from `<SOFTWARE_INSTALLDIR>\examples\tutorials\` and hit **Open.**
2. Carrying out a Ring search
   - Select all six atoms that make up the piperidine ring (N27 C30 C31 C32 C33 C34) and click **Search.**
3. Analysing the results
   - The histogram showing the distribution of geometries for related rings should appear. The X-axis gives an average measure of Root Mean Square Difference (RMSD) between each ring torsion in the query and the corresponding ring torsion in each hit.



- The search retrieves sixteen hits with relevance 1.0. All are at the far end of the X-axis at RMSD of 70 degrees indicating that the piperidine geometry in the model is very different from that represented in the CSD.
- Click on **More hits ...**, to bring up the *Find more hits [ring]* window, change the threshold in the **Aim for at least** box to 45, and click on **OK**. Does the distribution of ring geometries significantly change?

4. Finding more information: Analysing the torsions.

- We will now examine whether other geometrical features look odd. Go back to the *Build query* window and select **All fragments**.
- Toggle on **All torsion fragments** and ensure all the other options re toggled off. Click on **Search**.
- The *All fragments: Results* window that comes up has a column labelled *d(min)*. This measures in degrees the distance of the query torsion to the nearest occupied bin on the histogram. Click on the top of the *d(min)* column to sort it so that the highest *d(min)* value is at the top.



- Examine on the query structure, the torsions represented by the second to fourth rows. We will ignore the first row (can you think of the reason why we should do that?) Two torsions are represented, both close to the piperidine fragment.
- Look also at the histograms for these torsions. Do you think that the query values are reasonable ?
- We ignored the first row simply because there was only one example to compare with in the CSD. We would need more before we could decide this torsuion was unusual. If you like try carrying out a more generalised search on this torsion to find more similar examples.

5. Looking at a second model structure in *1hak.*

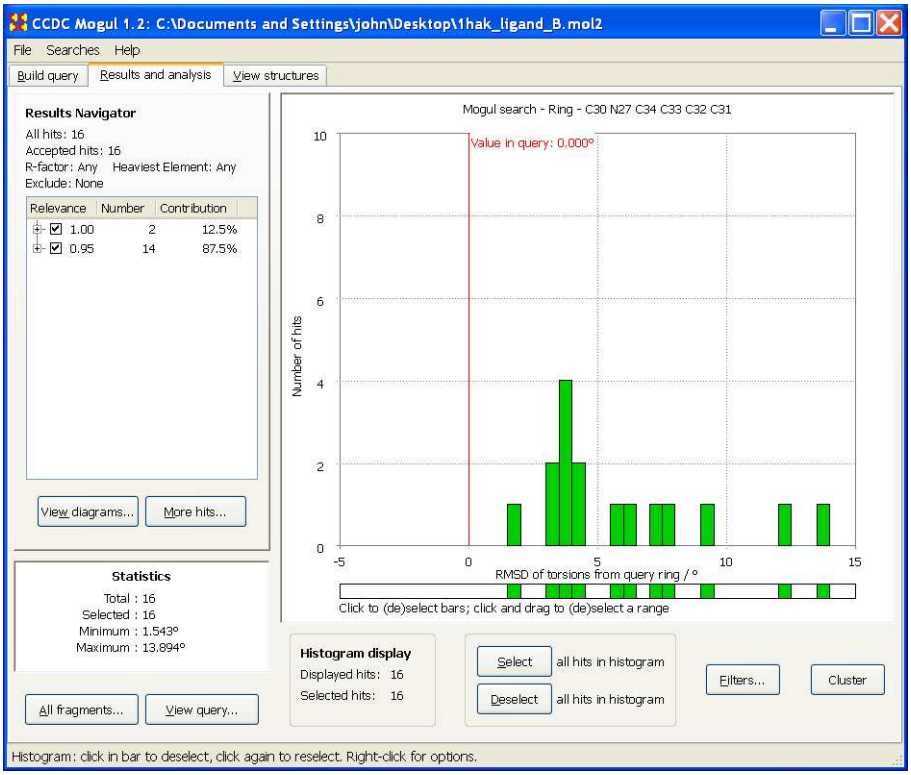
- The structure *1hak* has two molecules of protein in the unit cell. Consequently two model structures exist for the ligand bound in the active site. We would normally expect these to exhibit very similar binding mode and geometry although it cannot be ruled out that differences in geometry may really occur, due to real differences in the protein conformation brought about by crystallographic packing and environment. However in many cases a difference in geometry between two available ligand models is due to incorrect ligand model choice, in one or both cases
- Click on the **Load** button in the *Build query* pane. In the resulting *Load molecule* dialogue box, select `1hak_ligand_B.mol2` from `<SOFTWARE_INSTALLDIR>\examples\tutorials\` and hit **Open**

6. Carrying out a Ring search

- Select all six atoms that make up the piperidine ring (N27 C30 C31 C32 C33 C34) and click **Search.**

7. Analysing the results

- The histogram for the ring search indicates that the ring geometry is not particularly unusual. It is in fact close to a chair conformation, which is very common in six membered rings. Perhaps this indicates that this ligand model is a better one?

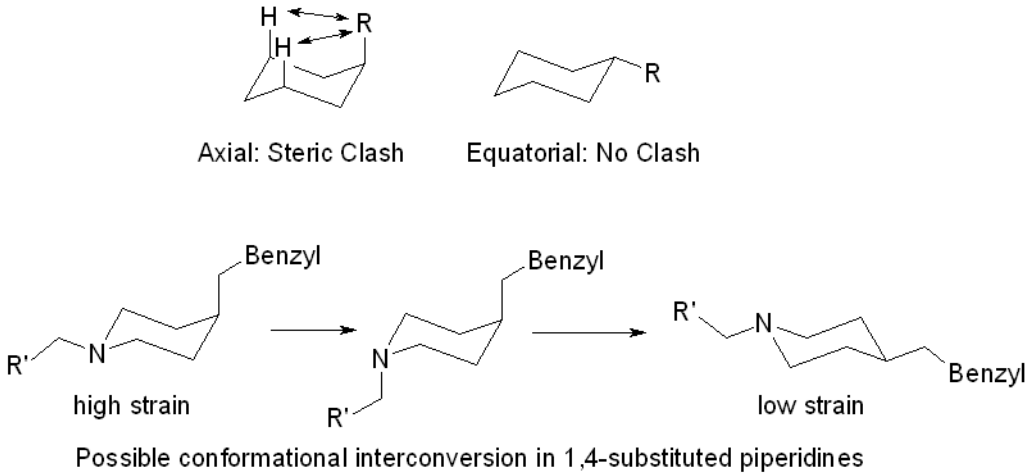

CCDC Mogul 1.2: C:\Documents and Settings\john\Desktop\1hak_ligand_B.mol2
File Searches Help
Build query Results and analysis View structures
Results Navigator
All hits: 16
Accepted hits: 16
R-factor: Any Heaviest Element: Any
Exclude: None
Relevance Number Contribution
1.00 2 12.5%
0.95 14 87.5%
View diagrams... More hits...
Statistics
Total : 16
Selected : 16
Minimum : 1.543°
Maximum : 13.894°
All fragments... View query...
Mogul search - Ring - C30 N27 C34 C33 C32 C31
Value in query: 0.000°
Number of hits
RMSD of torsions from query ring / °
Click to (de)select bars; click and drag to (de)select a range
Histogram display
Displayed hits: 16
Selected hits: 16
Select all hits in histogram
Deselect all hits in histogram
Filters...
Cluster
Histogram: click in bar to deselect, click again to reselect. Right-click for options.

- Try to find more hits. Very few hits of relevance 1.0 are found, which might be a significant observation.

8. Finding more information: Analysing the torsions.
   - We will again examine whether other geometrical features look odd. Go back to the *Build query* window and select **All fragments**.
   - Toggle on **All torsion fragments** and ensure all the other options are toggled off. Click on **Search**.
   - Again sort the *All fragments: Results* table by *d(min)*. Examine on the query structure the torsions represented by the top three rows. Two torsions are represented, both close to the piperidine fragment.
   - Look also at the histograms for these torsions. The query values are well away from the CSD distribution. Does this then mean that the model is bad after all?
9. A hypothesis
   - The `1hak_ligand_A.mol2` structure is very likely to be a bad model. The ring conformation is clearly highly unusual. What about the model `1hak_ligand_B.mol2` however ?
   - There is a clue to the answer in that very few examples of relevance 1.0 were found for model B. Relevance in rings depends on three factors, number of substituents on each ring atom, size (small or large) of each ring atom substituent; and relative stereochemistry. The piperidine ring is substituted only twice, in the 1 and 4 positions and the subtituents themselves are both considered 'small' (because the atom of each substituent adjacent to the ring only has at most

one additional heavy atom attached). It is unlikely that other similar examples cannot be found in the CSD, and in fact we know such examples exist because 9 examples with relevance 1.0 were found for the search on `1hak_ligand_A.mol2` . Therefore it must be the relative stereochemistry of the two ring substituents that is the unusual factor, and which forces the ring search to generalise and get other hits.

- Examination of the piperidine ring in `1hak_ligand_B.mol2` shows both substituents to be UP (or DOWN). Why is this so rare in a 1,4 substituted piperidine?
- The answer lies in the stereochemical preferences of substituted saturated six membered rings. So called *Axial* substitutions where the substituent points vertically up perpendicular to a plane through the ring (chair conformation assumed) are less stable (because they make more close contacts to ring H's) than *Equatorial* substituents which, as their name suggests, come off the sides of the ring (see diagram). The benzyl group in Model B coming off the a piperidine ring carbon, is in an *axial* position. The substituent off the nitrogen atom is in the *equatorial* position. This conformation for the piperidine ring is unusual because the nitrogen in similar compounds is usually able to invert itself (*umbrella inversion*) so that its substituent is *axial*. The six membered ring can then also invert by a process called *intra-chair conversion* and in this process both *axial* ring substituents can then take up the much more stable *equatorial* position. Consequently we see that the piperidine ring in Model B still contains significant strain, despite having a common chair conformation, and that this strain is at least in principle resolvable by adopting an alternative conformer.



Possible conformational interconversion in 1,4-substituted piperidines

- We see now that both model A and model B are highly strained structures. A full analysis would require us to look at the electron densities in both structures. However, even without that, a working hypothesis might be that the crystallographic refinement of both ligand models is not as good as it could be and that, if the crystallographer could have used alternative starting models for each ligand structure, it might have been possible to generate models with good fit to the

electron density, which contained low strain piperidine conformations, each a chair form with two equatorial substituents.

This ends the tutorial.